

Who'd want to work in a team?

Biologists and their institutions are increasingly confronted by the challenges of working in major collaborations that other disciplines have already addressed. A gathering last week showed how much further there is to go.

Team science is everywhere these days. The trouble is, you'd never guess it from an inspection of the universities that house it or the agencies that fund and supposedly foster it. Last week, a meeting at the US National Institutes of Health (NIH) on "Catalyzing team science" highlighted the difficulties, and proposed some solutions. What should disturb everybody is how far from reality many of those solutions are.

In one sense the challenges are nothing new. Particle and space physicists have been doing team science for decades. But many of those enterprises grew up around major facilities or dedicated institutions. Their goals were clear at the outset, and structures, processes and cultures developed accordingly. Furthermore, the disciplines involved — physics, astronomy, engineering and computation — are far from alien to one another.

Now fundamental biology and biomedical research are more and more facing similar challenges. But the teams are emerging within frameworks established for a more traditional ethos of investigator- and hypothesis-driven research. What is more, progress requires that biologists, chemists, physicists and engineers mingle and even merge their disciplinary cultures and languages — sometimes an extremely tall order.

Whether in genomics, chemical biology, nanotechnology or imaging for cancer research, the case studies at the NIH meeting time and again highlighted the need for good communication across the collaborations — weekly video meetings over the web seemed a minimal requirement for success. And personality is everything. "Pick people you can rely on," said one leader. "You cannot regulate for personality," said another, "but you can foster generosity of spirit."

Predictably, many of the recommendations were aimed at the NIH. And so they should be: the agency has identified team science as a key element of its 'road map', whereas, as the meeting frequently highlighted, its rules either obstruct team science or do too little to facilitate it. The meeting's conclusions (see www.becon.nih.gov/symposium2003.htm) now need to be followed through with the full support of those at the top in the NIH.

Essentials

Some golden rules emerged for team leaders. First, do not underestimate the demands of running the team. Even the simplest collaboration might require two principal investigators, two department chairs and two deans to reach agreements. Several relatively new major interdisciplinary centres highlight the importance of PhD administrators — scientists or ex-scientists who are responsible for ensuring that recruitment, grant applications and similar key tasks proceed in good order, while actively contributing to the development of the scientific programme. The lack of an adequate career track for such individuals and project leaders more generally was highlighted.

Another golden rule is to establish certain key principles at the outset of a collaboration. Take technology transfer. Not only should differences in disciplinary, institutional and even national practices be addressed in negotiations at the outset, but university technology-transfer departments should be closely involved from the start and

treated as part of the team, in order to minimize delays and obstacles further down the road. Bear in mind, said one participant, that technology-transfer revenues amount to only a few per cent of most institutions' income, and that there are very simple model agreements that can be applied in most cases, minimizing the need to reinvent wheels.

Another key aspect to be negotiated at the outset of a collaboration is the inescapable need for principles concerning team publications. Many collaborations progress splendidly, only to come to blows when it becomes necessary to list authors on a publication. The meeting highlighted how useless is the list of authors as a measure of their contributions, and urged journals to allow authors to publish lists of their respective contributions to a paper. *Nature* and its associated journals already do this, but will take further steps to encourage it, for example by including fields for such information in the electronic submission template. Indeed, many at the meeting urged that we and other journals should make such information compulsory. Only journal editors at the meeting expressed reservations on this. Readers are invited to send in their views.

Due recognition

But the big theme that emerged time and again related to the inability, because of the way science is currently done, to give credit and recognition to scientists who are part of a team. For example, too often, teams aim for a few high-profile publications that cover a lot of ground in a highly condensed manner. These miss opportunities not only to spell out interesting technical developments achieved along the way, but also to ensure that the people who delivered such innovations get any external recognition for them. Team leaders need to pay more attention to fostering additional publications in specialized journals, not as 'salami slices', but as appropriately focused accounts of genuinely innovative developments in techniques.

But the biggest challenge highlighted in the discussions is a scandal too often dismissed as insoluble: the way in which universities' appointments, promotion and tenure committees work or, more to the point, fail to work. The list of problems is a long one. Too often their considerations are overly systematized, so that consideration of candidates is inappropriately swayed by quantitative measures such as the impact factors of journals in which the candidate has published, the quantity of publications and whether or not the candidate was a first author. Too often the committees involve, say, history professors when deciding on the merits of a scientist whose multidisciplinary involvement makes him or her a challenge for even scientists to assess. Such a system is too clumsy to do justice to what should be crucial decisions for universities: the appointment of the people who will be their lifeblood. Yet, at the meeting at least, any idea that such committees could change their practices was dismissed out of court.

Given this and other difficulties in achieving due recognition, it is hard to see why anyone would want to become a team scientist. But as the meeting showed, funders, journals, universities and teams themselves can and should all take concrete steps to enhance their flexibility and to make the doing of team science as attractive to young researchers as the scientific challenges that necessitate it. ■





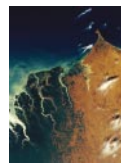
Ring out the old?
Torus reactor faces uncertain future amid closure threats
p4



Drug problem
Rivers and soils contaminated with waste antibiotics
p5



Power play
Japanese team bids to kick flagging probe into Mars orbit
p6



Eye in the sky
Satellites to keep tabs on UNESCO World Heritage sites
p8

Below-par performance hampers Fermilab quest for Higgs boson

Geoff Brumfiel, Washington

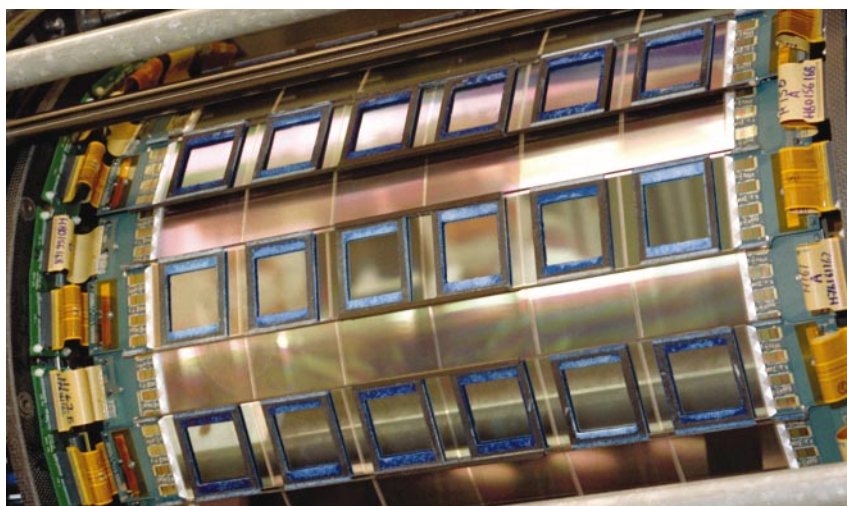
Physicists at the Tevatron particle accelerator near Chicago are steeling themselves for failure in their ambitious bid to detect the elusive Higgs boson.

Researchers working on the machine, at the Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, are searching for signs of the particle, thought to give other particles their mass, in the debris of high-energy particle collisions.

But they have now cut their estimate of the number of proton-antiproton collisions they expect to see by 2008 by 60–80%. As many collisions have to be studied to detect the Higgs boson, this is a serious blow to the lab's hopes of observing the particle.

The figures appear in a document prepared for Fermilab's sponsor, the US Department of Energy, and released on 15 June. Stephen Holmes, associate director for accelerators at the lab, says that problems have arisen with the equipment used to accelerate the protons and antiprotons. The Tevatron is 20 years old and its accelerators have been plagued by trouble during 'Run II', an upgraded second phase of operation that began in 2001.

The news is forcing Michael Witherell, Fermilab's director, to reconsider funding for Run II, which consumes nearly two-thirds of the lab's \$300-million annual budget. "We have to make some tough choices," he says.



Poor results mean that Fermilab may not splash out on new silicon wafers for its particle detectors.

Witherell says he may have to withhold \$25 million needed to replace the detectors' silicon wafers, which create electrical signals when hit by particles. Researchers warn that this could severely impair the detectors' performance.

Many observers doubt whether the Tevatron will be able to find the Higgs boson before the rival Large Hadron Collider comes online in 2007 at CERN, the European particle-physics lab near Geneva. Witherell puts the chance of spotting the Higgs at

"something like 50%". Others disagree: "I don't think there's any chance they will find it," says CERN physicist Daniel Froidevaux.

Many Fermilab researchers admit that they placed too much emphasis on finding the Higgs. Now, they say, they need to draw attention to their other research, such as studies of the top quark, a subatomic particle that was discovered at the lab in 1995. "I think we need to get some buttons out there that say: 'Run II, it ain't just the Higgs,'" says Holmes. ■

String theorists bypass NSF en route to Iran seminar

Geoff Brumfiel, Washington

Despite increasing political tension, a group of US researchers are standing by their pledge to visit Iran this autumn for a workshop on superstrings.

Lack of money almost forced the team to abandon the trip. Getting the funds has been difficult, says trip organizer Albion Lawrence of Brandeis University in Waltham, Massachusetts. Fundraising was delayed for months by the war in Iraq and the current row over Iran's nuclear programme.

But on 25 June the problem was solved for

seven members of the twelve-strong group. The Clay Mathematics Institute, a Boston-based charity, pledged \$12,000 for the researchers to travel to Tehran. The National Science Foundation (NSF) has denied funds to graduate students who want to attend.

The meeting is organized by the Iranian Institute for Studies in Theoretical Physics and Mathematics in Tehran, which has held an annual school and workshop on string theory for the past three years.

This year's conference, near Anzali, on the Caspian Sea, from 29 September to

9 October, will be the first to have significant US participation. "This is an important development in scientific relations between US and Iranian scientists," says Farhad Ardalan, a theorist at the Tehran institute and one of the meeting's organizers.

String theory postulates that particles can be described as vibrating loops or strings. "In the past few years, a strong group of string theorists has developed in Iran," says Eva Silverstein, a theorist at the Stanford Linear Accelerator Center in California who will attend the meeting. ■

Fusion cash shortfall leaves JET grounded

Jim Giles, London

One of the world's premier fusion experiments — the Joint European Torus (JET) reactor in Oxfordshire, UK — may have to shut down within the next few years, researchers and officials say.

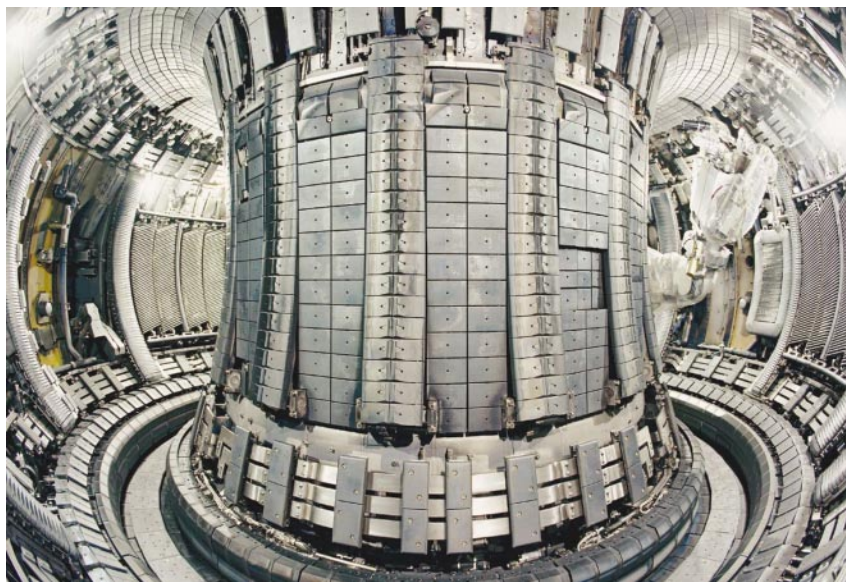
JET's future is in doubt because the current fusion budget at the European Union (EU) cannot both support the continent's present fusion projects and pay for the construction of ITER, the planned international magnetic fusion reactor.

ITER is expected to take about a decade to build. And closing JET such a long time before ITER is operational would severely disrupt preparations for the new reactor, critics of the closure proposal contend.

Some €200 million (US\$230 million) of the EU's €750-million fusion budget for 2002–06 has been earmarked for ITER's construction. If work begins before 2006, officials will have to cut funding to existing projects to make ends meet. Because JET consumes the single biggest chunk of EU fusion money — about €55 million per year — researchers fear that it may be selected for closure. "There is a possibility that JET will have to close by 2006," confirms Pascal Lallia, an energy adviser with the European Commission's research directorate.

Fusion researchers accept that cuts must be made to allow ITER to be built, but they are split on whether JET should be spared. Supporters point out that JET is the closest in size to ITER of Europe's fusion reactors, and is the ideal place for fusion scientists to work while ITER is being built. "JET could help develop a generation of fusion researchers to run ITER," says David Baldwin, head of fusion research at General Atomics in San Diego, California.

Others note that research at JET over the



Under threat: the Joint European Torus fusion reactor may be forced to shut down by 2006.

past decade has helped to streamline ITER's design. Future experiments at JET, to increase understanding of the turbulence in the plasma contained in the reactor and how to control it, for instance, are likely to help guide ITER's early work, they add. "It's in the collective interest of European fusion researchers to keep JET going," says Steven Cowley, a plasma physicist at Imperial College London.

But shutting JET might be simpler than cutting funds to several of the 12 other national fusion projects that receive money from the same EU funding stream, and which could also be used to train researchers for ITER. "In an ideal world we would keep JET going," says Alexander Bradshaw, scientific director of the Max Planck Institute for

Plasma Physics in Garching, Germany. "But if we are short of money for financing the EU contribution to ITER construction, it would be more sensible to shut it."

Negotiations about which facility to close will begin in earnest once a site has been chosen for ITER, which received a boost last week when South Korea announced that it would be joining the project. An expert group of scientists and engineers, led by David King, the UK government's chief scientific adviser, is assessing the merits of proposed sites in France and Spain and will report in September. ITER's international partners will then compare the European site with two others in Canada and Japan, and a final decision could be made by the end of this year. ■

Anger mounts over cutbacks at US army pathology lab

Erika Check, Washington

Scientists at a military medical facility in Washington DC have lashed out at plans by the Department of Defense (DOD) to cut back on both its staff and its unique mission.

Observers say that the cuts, at the Armed Forces Institute of Pathology (AFIP), would also affect thousands of scientists and clinicians in non-military hospitals, who benefit from the institute's expertise.

"The AFIP is an incredible resource," says Darlene Ketten, a biologist at Woods Hole Oceanographic Institute in Massachusetts. "Their pathologists are the only ones I can send tissues to and know

I will get an answer affirmed by everyone in the research community."

The 60-year-old AFIP has an annual budget of \$65 million and 785 staff, based on the campus of the Walter Reed Army Medical Center and at several smaller facilities in the Washington area. It investigates US military casualties, as well as deaths that occur on government land. It was scientists at AFIP, for instance, who identified 178 of the 184 people killed during the terrorist attack on the Pentagon on 11 September 2001.

But scientists also use the AFIP's archive of materials and records from almost eight million deaths. For example, Jeff

Taubenberger, a molecular pathologist at the AFIP, used genetic material from lung tissue dating from the First World War to sequence the virulence genes of the deadly Spanish influenza virus (A. H. Reid, T. G. Fanning, J. V. Hultin and J. K. Taubenberger *Proc. Natl Acad. Sci. USA* 96, 1651–1656; 1999).

The DOD began scrutinizing the AFIP five years ago, when the institute claimed that it needed half-a-billion dollars to replace ageing buildings and equipment. The department found that it was subsidizing the AFIP's non-military work and in 2001 issued a report recommending that the AFIP either impose severe cost-saving measures or transfer some of its work to the

Studies assess risks of drugs in water cycle

Quirin Schiermeier, Gothenburg

Antibiotics and other pharmaceuticals given to humans and livestock are increasingly contaminating rivers, groundwater and soils, according to early results from three European studies.

Drug-contaminated waste water is a potential risk to both human health and the environment, the studies' participants say. Treating sewage with ozone would be the best way of cutting the drug residues, they suggest.

Groups behind the three European Union-funded studies released their results at a Gothenburg press conference on 27 June. They reported that high concentrations of excreted antibiotics have been found in hospital and household sewage, slurry and water used for irrigation. Antibiotics and their metabolites also reach the environment directly from the urine and faeces of farm animals, the scientists find.

Researchers worry that the growing levels of antibiotics in the environment may damage ecosystems and fuel a surge in antibiotic

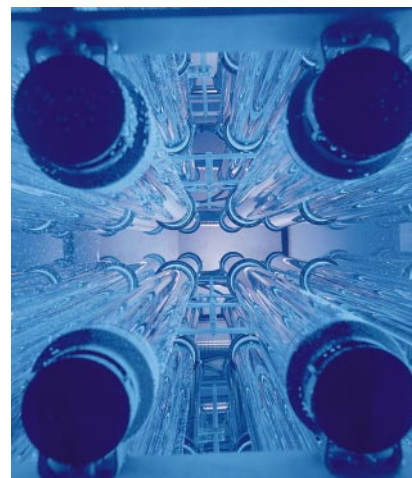
resistance among humans and animals. "Antibacterial resistance from farm animals can be transferred to humans, and is increasingly threatening the effective management of infectious diseases," says Wolfgang Witte, a microbiologist at the Robert Koch Institute in Wernigerode, Germany.

Officials at the European Union say the studies, and other similar ones, are likely to form the basis of new management procedures for medicines. Water companies and hospitals could also be required to take further steps to extract antibiotics from certain stages of the water cycle. This could be done, for example, by using separation techniques on the urine from hospital patients, and by injecting ozone into effluent at sewage plants.

Scientists involved in one of the projects, called Poseidon, found 35 pharmaceutical compounds, including five antibiotics, in the effluent of a local sewage plant in Braunschweig, Germany, and in local rivers. Concentrations in the effluent reached several micrograms per litre, with those in rivers and groundwater measuring 1.0–2.5 micrograms per litre, depending on the compound. Measurements taken at sewage plants in Poland, Switzerland, France and Austria revealed similarly high drug concentrations.

"Concentrations were particularly high in small rivers and streams that take a large amount of municipal waste water," says Thomas Ternes, a chemist at Germany's Federal Institute of Hydrology, and project coordinator of Poseidon.

In a parallel project, called ERAVMIS, researchers in Denmark, the Netherlands, Britain and Spain assessed the fate of veterinary medicines in slurry, made from manure, that is applied to the land as fertilizer.



Ozone treatment such as that used on drinking water (above) could strip drugs from sewage.

They found that tetracyclines, a group of antibiotics widely used in livestock, cling to soil, degrading slowly. Concentrations remained high six months after application.

A third project, carried out by researchers in Sweden, Greece, Italy and France, also monitored effluents from sewage plants.

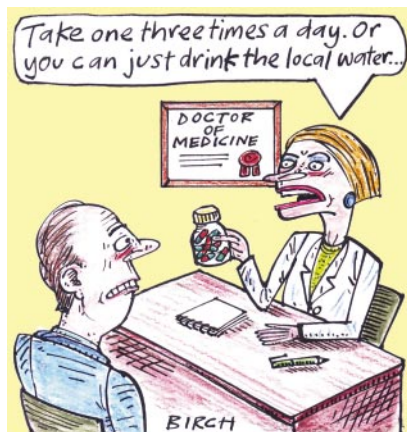
"There is clear evidence that some veterinary antibiotics persist in soils, and can move to rivers and streams and enter groundwater," says Alistair Boxall, an environmental chemist at Cranfield University, UK, and project coordinator of ERAVMIS.

The environmental effects of these compounds remain unclear, however. Researchers found no acute toxic effects on crops, worms, insects or fish. But they did find adverse effects on soil bacteria and some aquatic plants.

"Antibiotics can change the growth, enzyme activity and diversity of soil microbes," says Boxall. This may lead to a larger ecological problem, he adds, "but we don't know how big it is". More research is needed to determine whether microbial pathogens are becoming resistant to standard antibiotics, he says.

One recent analysis of raw sewage at the University Hospital of Würzburg, Germany, found that concentrations of antibiotics were not high enough to allow the selection of pathogenic bacteria resistant to them (K. Ohlsen *et al. Environ. Microbiol.*; in the press). But risks cannot be discounted on the basis of this result, says study co-author Thomas Ternes, a chemist at Germany's Federal Institute of Hydrology in Koblenz.

Ternes, the project coordinator of Poseidon, suggests that 'ozonation' of sewage is the most efficient way to eliminate unwanted residues. Ozone can be readily extracted from air, and it breaks down 99% of antibiotics and other drugs, he says.



Department of Health and Human Services.

In response, AFIP leaders told their staff this spring that they plan to save \$6.6 million a year by eliminating more than 100 jobs, and to slash the institute's \$3.5-million rent bill by closing three of its five facilities. They also decided to raise consultation and training fees, so that by 2004 the AFIP's military responsibilities would outweigh its civilian workload.

But AFIP staff claim that they were not consulted on the proposed changes, and many are strongly opposed to them. "I understand they had to make cuts but the way they proposed to do it was just malicious," says Renu Virmani, chair of

cardiovascular pathology at the institute. Critics add that the plan will damage the AFIP's ability to retain expert pathologists by restricting its activities in non-military research.

Renata Greenspan, a pathologist at the Walter Reed Army Institute of Research who became director of AFIP in May, is re-evaluating the business plan, and says that she has worked hard to include the institute's staff in the process. She says that she has not determined what combination of staff reduction, rent cuts and increased charges will be necessary, and expects to present a new plan to the AFIP's divisional heads at the end of this month.

Tribunal clears obesity researcher of fraud

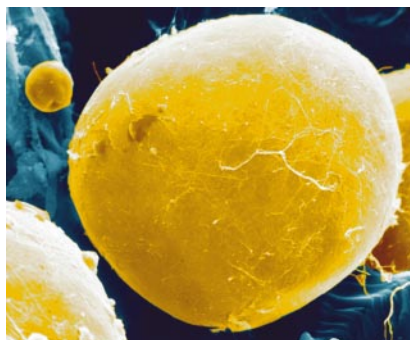
Declan Butler, Paris

A French tribunal has ruled that there are no grounds for the prosecution of Bernard Bihain on charges of 'forgery and use of forgeries'. Bihain, an obesity researcher, quit French public research in 1998 amid allegations of scientific misconduct.

Bihain's work on the identification and cloning of a molecule involved in fat degradation was queried in 1997 by members of his lab at the University of Rennes 1. An internal inquiry commissioned by the French research ministry concluded that the testimony of several whistle-blowers raised doubts about certain results produced by the lab (see *Nature* **391**, 519–520; 1998).

But an independent international scientific inquiry promised by the science ministry was dropped in October 1998 after Bihain closed his lab (see *Nature* **393**, 203; 1998 and *Nature* **395**, 829; 1998). Bihain went on to become vice-president of physiological genomics at Genset, a French biotechnology company, and later chief scientific officer of the US–French company ValiGen.

The new ruling, by the Tribunal de Grande Instance in Rennes, concludes a separate judicial investigation of some aspects of the affair. It was launched in December 1998, when the public prosecutor brought charges on the basis of a preliminary inquiry by the fraud squad which raised the question of "an alteration of the truth in the translation of the results of experiments ... to make them coincide with a hypothesis



Questions over research on fat cells (pictured) sparked accusations of scientific misconduct.

developed by Mr Bihain" (see *Nature* **391**, 825; 1998 and *Nature* **394**, 308; 1998).

The judicial investigation heard testimony from several researchers, who had made allegations of misconduct. In addition other testimony questioned the experimental competence of several of the whistle-blowers, and Bihain himself accused them of making false accusations.

In reaching its decision on the charges, the tribunal was only permitted by law to consider certain limited classes of documents. This is because the misdemeanour 'faux et usage de faux' ('forgery and use of forgeries') applies only to a "document or other medium of expression of which the object is, or effect may be, to provide evidence of a right or of a situation carrying legal consequences". This is not to say, how-

ever, that the tribunal did not look at other documents to understand the wider context.

The court concluded that the allegations of "scientific fraud" generally were "essentially a controversy of a scientific nature" — arguments among experts over methods, data interpretation and other elements of "ordinary experimental debate".

Michel Philippe, a researcher at the publicly funded laboratory of developmental biology and genetics at the University of Rennes 1, who took part in an earlier internal inquiry by the university, says he regrets that the promised independent, international scientific inquiry into the case was not conducted. INSERM, the French national biomedical agency, "has been unable to set up its own expert committee, and is now trusting a decision that was not made by scientists", he says.

But in a statement, INSERM said that it is delighted that the ruling has "put a definitive end to this affair and returned the honour to one of its researchers". Christian Bréchet, the agency's director general, says that the decision "honours the scientific integrity and the respect of the ethical rules common to all the researchers of the agency".

Bihain returned to INSERM last year as a researcher at the Nancy Centre of Clinical Investigation (see *Nature* **418**, 467; 2002). Bréchet personally handled Bihain's re-appointment, and INSERM's unit of scientific integrity, which was set up in the aftermath of the Bihain affair, was not consulted. Bréchet points out that he asked an international team of experts to review Bihain's past research and future plans, and that the results were generally positive. Bihain's future research will be subject to the agency's usual regular evaluation procedures, Bréchet adds.

Bernard Bigot, director of the cabinet of science minister Claudie Haigneré, told *Nature* that the affair is now closed, and that the ministry has no intention of revisiting it. Bigot is familiar with the episode: in 1997 he was research director at the ministry and set up the initial internal inquiry that received testimony from whistle-blowers. Bigot subsequently carried out his own inquiry into the affair, and in December 1997 he submitted a report to the ministry, clearing Bihain's laboratory of misconduct. "The conclusion is the same as it was five years ago," says Bihain: "that there was no scientific fraud."

Bihain, who claims to have been the victim of vendettas at the University of Rennes 1, hopes that the ruling will put an end to any suspicion that people may have held about him. The allegations and media coverage of the case "destroyed my credibility at the international level", he says. "I have continued to present my results to my peers while being an 'accused' researcher, whereas I never cheated in anything."

Astronomers try to save Mars probe

David Cyranoski, Tokyo

Japanese scientists will this month begin a last-ditch effort to save the Nozomi Mars probe, the country's first planetary mission.

The US\$848-million Nozomi was designed by Japan's Institute of Space and Astronautical Science (ISAS) to investigate the red planet's atmosphere and its interaction with the solar wind (see *Nature* **423**, 473; 2003).

But Nozomi ran into trouble soon after its July 1998 launch. A swing by Earth in December 1998 — intended to accelerate the probe using the planet's gravitational pull — failed to give it enough speed to reach Mars. Attempts to correct its course by firing its engines drained its fuel tanks. And during another Earth swing-by in April 2002, a solar flare damaged its power-supply system.

The flare seems to have damaged a circuit breaker that regulates the flow of electricity to some of the craft's essential systems,

including its heater. The breaker now "works too much, turning off and on continuously", says Yasunori Matogawa, ISAS scientist and director of Kagoshima Space Station from where Nozomi was launched.

Following a third, apparently successful, swing around Earth last week, the probe should finally reach Mars next January. But without the heater, the spacecraft cannot thaw the fuel it has left to manoeuvre itself into orbit around the red planet.

Later this month, ISAS scientists will try to fix the problem, sending successive commands to turn the power supply on. They hope that this will eventually burn out the circuit breaker, leaving the power switched on.

Matogawa says that they will probably know within the next couple of weeks whether the mission is likely to be saved. But even if initial efforts fail, they will keep trying until November. "I'd say there's a 50/50 chance," Matogawa says. ■

Collision debris yields five-quark particle

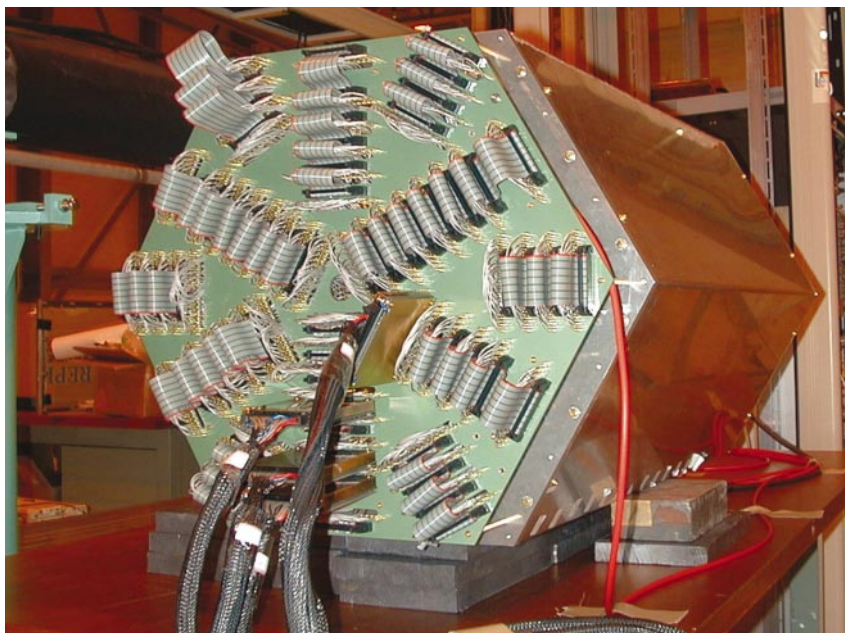
David Cyranoski, Tokyo

A class of subatomic particle, consisting of five quarks rather than the normal two or three, has been discovered by physicists in Japan. Theorists had expected combinations of four or more to exist, but experiments over the past 30 years had failed to detect them.

A team from the SPring-8 synchrotron in Harima, near Kobe, created the particles by firing γ -rays at a carbon target. The researchers, led by Takashi Nakano of Osaka University's Research Center for Nuclear Physics, say that the five-quark particle probably formed when two particles — a neutron, which is made up of three quarks, and a K^+ meson, which contains two — fused during the collision. Evidence of the particle, which decayed rapidly, was found by analysing debris from the collision.

The discovery, which is due to be published on 4 July in *Physical Review Letters*, will give a huge boost to particle-physics research, and could have important implications for our understanding of the early Universe.

Nakano himself might never have thought to look for the five-quark particle had it not been for a series of lunchtime conversations with Dmitri Diakonov of the Nordic Institute for Theoretical Physics in Copenhagen, while the pair attended a conference in 2000. Diakonov thought that the experiments that Nakano was planning could produce a five-quark particle, and suggested a new data-analysis method that he argued would reveal the particle's presence. "Dmitri suggested that the particle could exist as a smaller particle at a lower energy level, so that's where we looked," says Nakano.



The time-projection chamber at SPring-8, which played a role in the quest to find a five-quark particle.

"It was a novel way of interpreting the data."

Nakano presented his findings last October at the International Conference on Particles and Nuclei in Osaka. "We got a pretty negative response," he says. "Most people thought it was impossible." But in the months after the conference, a group at the Thomas Jefferson National Accelerator Facility in Newport News, Virginia, confirmed Nakano's results by revisiting data from previous experiments. The group presented its data, now submitted for publication, at the Conference on the Intersection

of Nuclear and Particle Physics, held in New York in May.

Nakano says that the discovery could help to pin down some theories of the early Universe — such as how quarks came together to form matter after the Big Bang. "Studies on two- and three-quark matter had hit a wall," he says. "This might provide more clues."

But he adds that the five-quark particles are not likely to be found in many places in today's Universe. "It's very unlikely that these exist anywhere freely, except maybe in a black hole," says Nakano. ■

Oxford professor accused of discrimination over e-mail

Haim Watzman, Jerusalem

A pathologist at the University of Oxford is facing an internal inquiry after he refused an Israeli student's PhD application because the student had served in Israel's army.

The student, Amit Duvshani, is currently studying in the medical faculty of Tel Aviv University. He sent a letter and curriculum vitae to Andrew Wilkie of Oxford's Weatherall Institute of Molecular Medicine because he was interested in Wilkie's research field of birth defects in embryos.

Wilkie responded by e-mail, advising Duvshani in no uncertain terms to look elsewhere. "I have a huge problem with the way that the Israelis take the moral high ground from their appalling treatment in the Holocaust, and then inflict gross

human rights abuses on the Palestinians because they (the Palestinians) wish to live in their own country," he wrote, adding: "I am sure that you are perfectly nice at a personal level, but no way would I take on somebody who had served in the Israeli army."

Duvshani shared the e-mail among students and faculty members in the medical school, who in turn circulated it more widely to scientists and news organizers — prompting outrage at what many see as overt discrimination against a student from a country where most young people are conscripted to serve in the military. Several angry protests were made directly to Wilkie, and to the University of Oxford.

Wilkie has issued an apology, which

was posted on the university's website on 27 June. "I recognise and apologise for any distress caused by my e-mail of 23 June and the wholly inappropriate expression of my personal opinions in that document," he wrote. Duvshani says that Wilkie subsequently contacted him personally, seeking his reaction to the incident.

The University of Oxford issued a statement announcing an internal investigation, which a spokesman says will be completed this week. "Freedom of expression is a fundamental tenet of university life," the statement said, "but under no circumstances are we prepared to accept or condone conduct that appears to, or does, discriminate against anyone on grounds of ethnicity or nationality, whether directly or indirectly." ■

Japan joins Europe's bid to get Bepi Colombo to Mercury

Tokyo Europe's plans to send a satellite to little-visited Mercury received a boost last week when the Japanese government signalled its intention to invest in the European Space Agency's (ESA's) Bepi Colombo mission.

Japan's education ministry, which is responsible for the country's space programme, provisionally approved ¥13.5 billion (US\$110 million) in funding for Bepi Colombo on 24 June. The mission — named after Italian physicist Giuseppe Colombo, who calculated the orbital manoeuvre that enabled NASA's Mariner 10 to pass by Mercury in 1974 and 1975 — is expected to launch early next decade. It will include the first lander to touch down on Mercury's surface. The mission, in which ESA is to invest €600 million (US\$690 million), will also study the planet's magnetic field and atmosphere, and will map its surface.

Mariner 10, the only previous spacecraft to travel to Mercury collected data on part of the planet. A second NASA mission, Messenger, is scheduled to launch in 2004. Final approval for the Japanese funding is expected to be granted next month.

Heads in the clouds

London A pair of British balloonists went on stand-by this week for an attempt on the world altitude record for a manned balloon. The QinetiQ 1 team will use the largest balloon ever made — as tall as the Empire State Building.

Andy Elson and Colin Prescott aim to reach 40 km — 5 km higher than the current record — before splashing down the same day in the sea off Cornwall in southwest England. The pilots will sit on an open platform, wearing spacesuits to protect against the harsh temperatures and space-like vacuum of the high stratosphere.

While one controls the balloon (shown here dwarfing Nelson's column), the other will attempt to capture the bid on video by steering a solar-powered aeroplane tethered to the gondola. Funded by QinetiQ, the corporate arm of the government's defence research agency, the team will also monitor the Sun's activity, as a solar flare during the ascent could expose them to lethal doses of radiation.



QINETIQ 1

Young women astronomers are in the ascendant

Washington More than half of the youngest astronomers in the United States are women, according to a new survey conducted by the American Astronomical Society (AAS).

The survey shows that roughly 57% of the society's members born between 1980 and 1985 are female. The next age group, born between 1975 and 1980, contains about 40%

women, but women make up less than 10% of senior members of the profession.

The number of female researchers in this field is notably higher than in the rest of the physical sciences. According to a 2002 study by the National Science Foundation, women make up only about 23% of the US physical-sciences workforce.

The new survey was released on 27 June at a conference on women in astronomy, hosted by the AAS in Pasadena, California.

Repository to nurture standard stem cells

Washington A repository for adult stem-cell research could soon be helping scientists working in this fledgling field.

The US National Center for Research Resources (NCRR), part of the National Institutes of Health (NIH), announced on 24 June that it will provide \$4.3 million over five years for the centre.

Attempts to investigate adult stem cells are currently held back by a lack of standardized cell lines. The repository, to be based at Tulane University in New Orleans, Louisiana, aims to solve this by producing and distributing human and rat adult stem cells, more properly known as mesenchymal stem cells. The cells can differentiate into many tissues but — unlike embryonic stem cells — apparently not all tissue types.

German foundation calls for African research partners

Munich The German Volkswagen Foundation has launched what it calls a unique funding programme to support research in sub-Saharan Africa.

Funding will be considered for any important research area with or without a direct African connection, excluding

developmental or historical themes. “We won’t be imposing research ideas on Africa,” says Antje Gunsenheimer, a spokeswoman for the Hanover-based charity. “And we want the researchers to work in Africa, not Germany.” An important focus will be helping young scientists to establish research programmes in countries that have not attracted Western money in the past.

In its first phase, €3 million (US\$3.4 million) will be available for projects each year, although this may increase. Ideas will emerge from a series of workshops on themes suggested by African researchers. Joint African–German proposals will then be assessed through the foundation’s normal peer-review procedure.

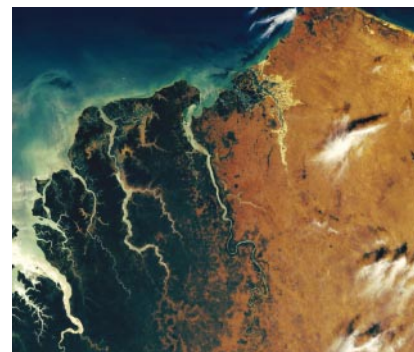
Eyes in the sky to look out for heritage sites

Paris World Heritage sites are to get a pair of orbiting guardian angels. The European Space Agency (ESA) announced last week that its remote sensing satellites will monitor the 730 sites and alert authorities to any threats.

The list of sites, which is maintained by the United Nations Educational, Scientific and Cultural Organization (UNESCO), includes many in developing countries that have not been extensively mapped. ESA and

UNESCO have already begun a collaboration to protect some heritage sites in the remote mountain areas of Uganda, Rwanda and the Democratic Republic of Congo, where hunting and farming are threatening gorilla populations. Radar images from ESA’s Envisat or ERS spacecraft will be used to alert authorities to changes in land use that could threaten the area.

It is hoped that other space agencies will join the partnership. ESA says that NASA is likely to sign up in October to mark its return to UNESCO after a 13-year absence (see *Nature* **419**, 235; 2002) and that the Indian, Japanese, Canadian and Brazilian space agencies have also expressed interest.



Erode map: Envisat shows how sediment pours down Senegal’s rivers and into the Atlantic Ocean.

ESA

Raiding the medicine cabinet

Many candidate drugs to fight diseases in the developing world have been shelved before approval — until now. Enter a medical charity that made its name by bringing hope to the world's disaster zones. Declan Butler reports.



Bucking the trend: Médecins Sans Frontières aims to beat the neglected diseases of the developing world.

When James Orbinski, then president of Médecins Sans Frontières (MSF), received the 1999 Nobel Prize for Peace on behalf of the charity, his acceptance speech amounted to a call to arms. "More than 90% of all death and suffering from infectious diseases occurs in the developing world," Orbinski told the assembled great and good. "Some of the reasons are that life-saving essential medicines are either too expensive or are not available because they are not seen as financially viable, or because there is virtually no new research and development for priority tropical diseases. This market failure is our next challenge."

This week, Paris-based MSF makes good on that pledge, by teaming up with a consortium of other organizations to launch a non-profit enterprise dedicated to developing essential drugs and distributing them to the world's poorest. For no more than US\$26 million annually, the Drugs for Neglected Diseases initiative (DNDi) aims over the next decade to register six or seven new drugs and to have another eight in the pipeline.

MSF — or Doctors Without Borders — won the Nobel peace prize for putting itself in the firing line, literally. In combat zones judged too dangerous by most aid agencies, its staff have held their ground and delivered medical care to some of the world's most desperate people. No one, therefore, doubts the commitment and courage of MSF's staff. But reinventing the economics of the drug industry presents a different challenge. Can

MSF and its partners really succeed where the pharmaceutical giants, and their multi-billion-dollar budgets, have failed?

Yes, claims Bernard Pecoul, who heads MSF's Access to Essential Medicines campaign. "It's an alternative model based on user needs and equitable access rather than profit," he says.

The DNDi's backers argue that drug firms are failing to address many diseases in developing countries because there is no commercial market. In the first instance, the initiative will concentrate on three 'orphan' diseases caused by single-celled parasites: African sleeping sickness, its Latin American relative Chagas' disease and leishmaniasis. Although malaria and tuberculosis are bigger killers, say DNDi officials, public-private partnerships are already striving to develop new treatments for these diseases.

Controlled approach

"We would like to see more diseases covered in the future, but you cannot deal with too many things at the same time," says José Roberto Ferreira, director of international relations with the Brazilian state-owned drug manufacturer Fiocruz, one of the DNDi partners.

In addition to MSF and Fiocruz, the initiative is backed by the Special Programme for Research and Training in Tropical Diseases (TDR) — a Geneva-based initiative sponsored by the United Nations, the World Bank and the World Health Organization — and by the Pasteur Institute in Paris, the Indian

Council of Medical Research, the Malaysian health ministry, and the Kenya Medical Research Institute (KEMRI) in Nairobi. Each partner will provide cash or research, and each has considerable experience either with the target diseases or in bringing drugs to the poor at affordable prices. "We know what we are talking about," says Davy Koech, director of KEMRI, which is a world-leader in sleeping-sickness research. Fiocruz, meanwhile, has successfully battled the US government and drug industry to win the right to make cheap copies of patented HIV drugs.

There is certainly a pressing need for new drugs to treat the diseases prioritized by the DNDi. Imagine the outcry in North America or Europe if a drug for, say, congestive heart disease didn't work in one-third of patients, and killed up to 10% of those who took it. Yet those are the shocking statistics for a 50-year-old drug called melarsoprol, used to treat sleeping sickness. The appallingly high rate of fatal reactions is tolerated because the disease otherwise represents a death sentence. But surely it isn't beyond the expertise of today's researchers to develop drugs that are both safer and more effective?

The problem is that the industry's drug-discovery work for such diseases has ground to a halt. Nonetheless, there is hope. DNDi officials say that the R&D portfolios of big drug firms contain candidate drugs for sleeping sickness, and other scourges of the developing world, that never became final products. So to kick off, the DNDi will attempt cheap and fast projects, taking no more than

R. JOB/MSF



The Kenya Medical Research Institute has extensive expertise in tropical parasitic diseases.

six years, to do the extra research to push such drugs over the regulatory hurdles needed to bring them to market. It will also examine whether the drugs might work better if they are reformulated, or used in different combinations. As an alternative to melarsoprol, for instance, the DNDi is considering a combination of meglazol, first synthesized in 1968, but never put through clinical trials, and nifurtimox, which is used to treat Chagas' disease.

Over the past year, the DNDi's backers have e-mailed and telephoned across academia and industry, and organized brainstorming sessions worldwide. This effort, combined with an extensive literature search, has turned up 15 lead compounds that might be used to treat African sleeping sickness, 11 for Chagas' disease, and seven for leishmaniasis, says

Pecoul. Some of these leads will be selected for more costly and risky projects, operating over a decade or so, to develop completely new drugs. In March, the DNDi and TDR also launched a call for research proposals, generating 71 responses. About half a dozen will be approved in September by the DNDi's scientific advisory committee, to be appointed at a meeting in Geneva on 3 July.

Good will hunting

How achievable are the DNDi's goals? An annual budget of US\$26 million is peanuts compared with the billions that companies spend on developing drugs for the rich world's diseases, and raising even that sum will require considerable effort. MSF plans to commit up to US\$7.5 million per year of its own funds to the DNDi over the next five years, and will engage its well-honed public-relations machine to raise money.

DNDi director Yves Champey, a former vice-president of R&D at the French drug firm Rhône-Poulenc Rorer, says that the budget will go far, because a lot of the help will be in kind. Champey is working for free, drug companies will give free access to their libraries of chemicals, and academic labs will dovetail projects funded from other sources to fit with the DNDi's goals. Clinical trials, usually the most expensive part of bringing a drug to market, will be performed cheaply at centres in developing countries where the diseases are endemic, Champey points out.

Champey's upbeat view is shared by the French arm of the business consultancy

Ernst & Young, which helped to draft a feasibility study and business plan for the initiative. "The people around the DNDi table are exceptional," says Thomas Saugnac, a consultant for the firm. "If this group can't pull it off, then who can?"

But promising drug leads often fall by the wayside before reaching the clinic because of toxicity, or because they don't work as well as was hoped. If the DNDi backs too many losers, its budget will soon disappear. With little margin for error, acting ruthlessly to kill projects that are going nowhere will be key to survival. And that will mean attracting individuals with relevant industry experience — no easy task, given that the DNDi does not offer fat-cat salaries.

Other experts worry that the DNDi's distributed network of partners could degenerate into a bureaucratic quagmire. "With such a disparate group of independent people and many divergent interests, I would predict that the feasibility of DNDi will not be very high," cautions Paul Herrling, head of corporate research at the drug firm Novartis in Geneva. There are also complaints from some quarters that the DNDi has not sought to bring in the drug giants and biomedical research agencies from rich industrialized countries as full partners.

Champey defends the initial choice of developing-country partners, but says that the DNDi intends to invite other organizations on board as the initiative gathers steam. Some observers wonder whether MSF's ethos may make this difficult. Used to operating under very difficult conditions, MSF has something of a guerrilla mindset and has been highly critical of drug companies. "This attitude is unnecessarily antagonistic, and could be counterproductive," warns Tony Holder, head of parasitology at the UK Medical Research Council's National Institute for Medical Research in London. The DNDi should be working more closely with drug companies, he argues.

Champey concedes that MSF has been reluctant to involve industry directly, but says that the DNDi is counting on such companies taking part and has already had fruitful private discussions with major players including Jean-Pierre Garnier, chief executive of Glaxo-SmithKline. Danielle Drew, a spokeswoman for the company, confirms that the DNDi is welcome to review GlaxoSmithKline's library of compounds "to identify any projects or molecules that were not progressed for technical or commercial reasons".

For every sceptic, another expert is willing MSF and its partners to succeed. "It's great that organizations such as MSF, who are confronted with diseases in very difficult countries, should take this up," says Ok Pan-nenborg, senior adviser for health nutrition and population at the World Bank.

Declan Butler is Nature's European correspondent.

♦ www.accessmed-msf.org/dnd/dndi.asp



Bitten: a telltale lesion shows that this Ugandan teenager has been infected with sleeping sickness.

Does the proof stack up?

Think peer review takes too long? One mathematician has waited four years to have his paper refereed, only to hear that the exhausted reviewers can't be certain whether his proof is correct. George Szpiro investigates.



Grocers the world over know the most efficient way to stack spheres — but a mathematical proof for the method has brought reviewers to their knees.

Just under five years ago, Thomas Hales made a startling claim. In an e-mail he sent to dozens of mathematicians, Hales declared that he had used a series of computers to prove an idea that has evaded certain confirmation for 400 years. The subject of his message was Kepler's conjecture, proposed by the German astronomer Johannes Kepler, which states that the densest arrangement of spheres is one in which they are stacked in a pyramid — much the same way as grocers arrange oranges.

Soon after Hales made his announcement, reports of the breakthrough appeared on the front pages of newspapers around the world. But today, Hales's proof remains in limbo. It has been submitted to the prestigious *Annals of Mathematics*, but is yet to appear in print. Those charged with checking it say that they believe the proof is correct, but are so exhausted with the verification process that they cannot definitively rule out any errors. So when Hales's manuscript finally does appear in the *Annals*, probably during the next year, it will carry an unusual editorial note — a statement that parts of the paper have proved impossible to check.

At the heart of this bizarre tale is the use of computers in mathematics, an issue that has split the field. Sometimes described as a 'brute force' approach, computer-aided

proofs often involve calculating thousands of possible outcomes to a problem in order to produce the final solution. Many mathematicians dislike this method, arguing that it is inelegant. Others criticize it for not offering any insight into the problem under consideration. In 1977, for example, a computer-aided proof was published for the four-colour theorem, which states that no more than four colours are needed to fill in a map so that any two adjacent regions have different colours^{1,2}. No errors have been found in the proof, but some mathematicians continue to seek a solution using conventional methods.

Pile-driver

Hales, who started his proof at the University of Michigan in Ann Arbor before moving to the University of Pittsburgh, Pennsylvania, began by reducing the infinite number of possible stacking arrangements to 5,000 contenders. He then used computers to calculate the density of each arrangement. Doing so was more difficult than it sounds. The proof involved checking a series of mathematical inequalities using specially written computer code. In all, more than 100,000 inequalities were verified over a ten-year period.

Robert MacPherson, a mathematician at the Institute for Advanced Study in Princeton, New Jersey, and an editor of the *Annals*,

was intrigued when he heard about the proof. He wanted to ask Hales and his graduate student Sam Ferguson, who had assisted with the proof, to submit their finding for publication, but he was also uneasy about the computer-based nature of the work.

The *Annals* had, however, already accepted a shorter computer-aided proof — the paper, on a problem in topology, was published this March³. After sounding out his colleagues on the journal's editorial board, MacPherson asked Hales to submit his paper. Unusually, MacPherson assigned a dozen mathematicians to referee the proof — most journals tend to employ between one and three. The effort was led by Gábor Fejes Tóth of the Alfréd Rényi Institute of Mathematics in Budapest, Hungary, whose father, the mathematician László Fejes Tóth, had predicted in 1965 that computers would one day make a proof of Kepler's conjecture possible.

It was not enough for the referees to rerun Hales's code — they had to check whether the programs did the job that they were supposed to do. Inspecting all of the code and its inputs and outputs, which together take up three gigabytes of memory space, would have been impossible. So the referees limited themselves to consistency checks, a reconstruction of the thought processes behind each step of the proof, and then a

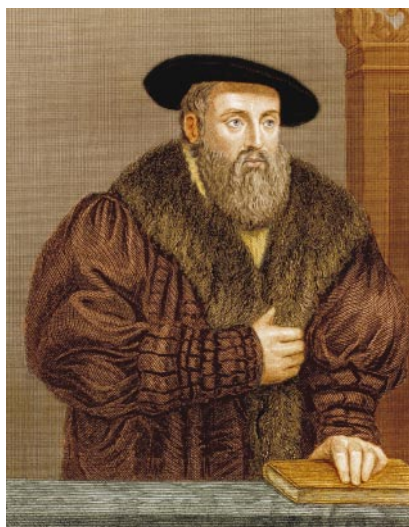
study of all of the assumptions and logic used to design the code. A series of seminars, which ran for full academic years, was organized to aid the effort.

But success remained elusive. Last July, Fejes Tóth reported that he and the other referees were 99% certain that the proof is sound. They found no errors or omissions, but felt that without checking every line of the code, they could not be absolutely certain that the proof is correct.

For a mathematical proof, this was not enough. After all, most mathematicians believe in the conjecture already—the proof is supposed to turn that belief into certainty. The history of Kepler's conjecture also gives reason for caution. In 1993, Wu-Yi Hsiang, then at the University of California, Berkeley, published a 100-page proof of the conjecture in the *International Journal of Mathematics*⁴. But shortly after publication, errors were found in parts of the proof. Although Hsiang stands by his paper, most mathematicians do not believe it is valid.

After the referees' reports had been considered, Hales says that he received the following letter from MacPherson: "The news from the referees is bad, from my perspective. They have not been able to certify the correctness of the proof, and will not be able to certify it in the future, because they have run out of energy ... One can speculate whether their process would have converged to a definitive answer had they had a more clear manuscript from the beginning, but this does not matter now."

Pyramid power:
Thomas Hales
believes that
computers will
succeed where
humans have failed
in verifying
his proof.



Star player: Johannes Kepler's conjecture has kept mathematicians guessing for 400 years.

The last sentence lets some irritation shine through. The proof that Hales delivered was by no means a polished piece. The 250-page manuscript consisted of five separate papers, each a sort of lab report that Hales and Ferguson filled out whenever the computer finished part of the proof. This unusual format made for difficult reading. To make matters worse, the notation and definitions also varied slightly between the papers.

Rough but ready

MacPherson had asked the authors to edit their manuscript. But Hales and Ferguson did not want to spend another year reworking their paper. "Tom could spend the rest of his career simplifying the proof," Ferguson said when they completed their paper. "That doesn't seem like an appropriate use of his time." Hales turned to other challenges, using traditional methods to solve the 2,000-year-old honeycomb conjecture, which states that of all conceivable tiles of equal area that can be used to cover a floor without leaving any gaps, hexagonal tiles have the shortest perimeter⁵. Ferguson left academia to take a job with the US Department of Defense.

Faced with exhausted referees, the editorial board of the *Annals* decided to publish the paper—but with a cautionary note. The paper will appear with an introduction by the editors stating that proofs of this type, which involve the use of computers to check a large number of mathematical statements, may be impossible to review in full. The matter might have ended there, but for Hales, having a note attached to his proof was not satisfactory.

This January, he launched the Flyspeck project, also known as the Formal Proof of Kepler. Rather than rely on human referees, Hales intends to use computers to verify

every step of his proof. The effort will require the collaboration of a core group of about ten volunteers, who will need to be qualified mathematicians and willing to donate the computer time on their machines. The team will write programs to deconstruct each step of the proof, line by line, into a set of axioms that are known to be correct. If every part of the code can be broken down into these axioms, the proof will finally be verified.

Those involved see the project as doing more than just validating Hales's proof. Sean McLaughlin, a graduate student at New York University, who studied under Hales and has used computer methods to solve other mathematical problems, has already volunteered. "It seems that checking computer-assisted proofs is almost impossible for humans," he says. "With luck, we will be able to show that problems of this size can be subjected to rigorous verification without the need for a referee process."

But not everyone shares McLaughlin's enthusiasm. Pierre Deligne, an algebraic geometer at the Institute for Advanced Study, is one of the many mathematicians who do not approve of computer-aided proofs. "I believe in a proof if I understand it," he says. For those who side with Deligne, using computers to remove human reviewers from the refereeing process is another step in the wrong direction.

Despite his reservations about the proof, MacPherson does not believe that mathematicians should cut themselves off from computers. Others go further. Freek Wiedijk, of the Catholic University of Nijmegen in the Netherlands, is a pioneer of the use of computers to verify proofs. He thinks that the process could become standard practice in mathematics. "People will look back at the turn of the twentieth century and say 'that is when it happened,'" Wiedijk says.

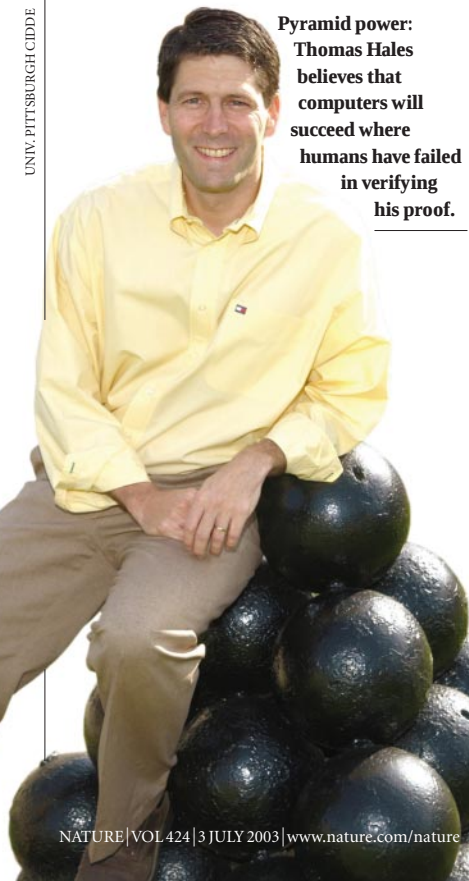
Whether or not computer-checking takes off, it is likely to be several years before Flyspeck produces a result. Hales and McLaughlin are the only confirmed participants, although others have expressed an interest. Hales estimates that the whole process, from crafting the code to running it, is likely to take 20 person-years of work. Only then will Kepler's conjecture become Kepler's theorem, and we will know for sure whether we have been stacking oranges correctly all these years. ■

George Szpiro writes for the Swiss newspapers NZZ and NZZ am Sonntag from Jerusalem, Israel. His book *Kepler's Conjecture* (Wiley, New York) was published in February.

1. Appel, K. & Haken, W. *Illinois J. Math.* **21**, 429–490 (1977).
2. Appel, K., Haken, W. & Koch, J. *Illinois J. Math.* **21**, 491–567 (1977).
3. Gabai, D., Meyerhoff, G. R. & Thurston, N. *Ann. Math.* **157**, 335–431 (2003).
4. Hsiang, W.-Y. *Int. J. Math.* **4**, 739–831 (1993).
5. Hales, T. C. *Discrete Comput. Geom.* **25**, 1–22 (2001).

Flyspeck

► www.math.pitt.edu/~thales/flyspeck/index.html



Impact factors: a tool of the sterile audit culture

Stalked by accountants, unaware of pre-Internet work, how can young scientists thrive?

Sir — I have added Peter A. Lawrence's Commentary on "The politics of publication" (*Nature* **422**, 259–261; 2003) and David Colquhoun's incisive Correspondence response "Challenging the tyranny of impact factors" (*Nature* **423**, 479; 2003) to a very small collection of items on this subject which are *de rigueur* reading for everyone in my group, no matter how senior or copiously cited. The two others are "Is science losing its objectivity?" (J. Ziman, *Nature* **382**, 751–754; 1996) and "The martial art of scientific publication" (E. N. Parker, *Eos* **78**, 393–395; 1997).

Asking the young scientists for whom I have responsibility to read these articles is the best way I can think of to meet Lawrence's question of what we can do about the situation that he so clearly diagnosed. The audit culture and the bragging attitude it sustains are advancing on all fronts; accountants and administrators armed with networked

computers are the downside of the advances that have transformed the way in which we are able to analyse scientific data. There seems to be no appreciation of the once unchallenged argument that the scientific goose that lays the golden eggs needs some tall green grass, privacy and free choice of nesting sites. It does not respond well to weekly or quarterly requests for deliverables and activity reports.

I fear that electronic publishing, while conferring obvious benefits such as connectivity and searchability of the literature, may exacerbate the situation. I have had a paper to review which cited about 30 references, hardly any of them more than 10 years old. When I commented that this was unfair to earlier workers, I received the response that older references were not available on the e-accessible database at the author's institution!

This anecdote typifies a detectable

trend and, I believe, is one that risks causing serious erosion of the knowledge base — not to mention the fact that the pioneers in a subject often stated the principles and problems with far greater clarity than many modern authors do.

One suggestion would be for a team of senior scientists, or perhaps national academies around the globe, to produce a handbook on responsibilities, rights and privileges as regards publication. It would be for all scientists but specifically aimed at entrants to the research enterprise. Even if one could not get agreement from those who successfully operate the present system to their advantage, at least the debate, if public, would raise consciousness.

Adrian Tuck

National Oceanic and Atmospheric Administration Aeronomy Laboratory R/AL6, 325 Broadway, Boulder, Colorado 80305-3328, USA

Reproductive cloning: don't rush to judgement

Sir — Your News story "Researchers divided over ethics of a ban on cloning" (*Nature* **423**, 373; 2003), describing the possible lack of ethical grounds for banning safe reproductive cloning, shows the perils of legislating early in the developmental trajectory of a technology. If human reproductive cloning becomes safe, protecting the welfare of cloned persons — who will be as unique and worthy of respect as any other person — will cease to be a strong reason to oppose reproductive cloning.

A more fruitful policy approach would then be to ask whether safe cloning would serve important reproductive or familial needs, and if so, what the impact of allowing cloning in those cases would be. An important distinction in this regard is between cloning to establish a family connection, as might occur in the case of severe gametic infertility, and cloning by fertile persons to choose the genotype of a child. Cloning when infertile to have an otherwise unavailable genetic connection with a child serves a different need and is arguably more deserving of societal respect than cloning by a fertile couple to choose a particular genome.

Whether legal bans are needed if cloning is safe should depend upon a much finer-grained policy analysis than has

occurred in the current rush to prohibit all cloning. If safe uses of cloning are not feasible, few responsible practitioners will offer the procedure. Even if cloning can be made safe, few otherwise fertile persons are likely to seek it or have a legitimate claim for it.

Legislating now to ban all cloning carries a high price, both in limiting potential future legitimate uses and in preventing researchers from cloning embryos for stem-cell or genetic-disease research. The possible dangers involved in reproductive cloning are too vague and unrealized to drive national and international policy covering all forms of cloning.

John A. Robertson

University of Texas School of Law, 727 East Dean Keeton Street, Austin, Texas 78705, USA

Reproductive cloning: an attack on human dignity

Sir — In your News story "Researchers divided over ethics of a ban on cloning" (*Nature* **423**, 373; 2003), about the German government's proposed ban on all types of human cloning, you report that some bioethicists regard the main argument against human cloning to be the risk of spontaneous abortion or ill-health in the offspring. Hence they argue that, if these

risks were overcome, other arguments against human cloning would not withstand ethical analysis.

However, I suspect that most people do not delegate their moral or ethical judgements to professional ethicists, but make judgements themselves on the basis of their personal philosophy (which may or may not have a religious underpinning) or of the 'gut feeling' that results from this.

I personally am completely opposed to human cloning of any sort, from a feeling of utter repugnance towards what appears to me to be a fundamental assault on human dignity, with the potential for horrendous misuse.

If this is a minority opinion among biological scientists, I suspect that it has a wider resonance among the general public, especially in countries with memories of the infamies perpetrated in the name of science and medicine.

I, like many scientists, am an atheist, but I do not think that anyone has a licence to play God.

David P. Leader

Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

The mind behind me

How and why does the brain generate a sense of self?

The Face in the Mirror: The Search for the Origins of Consciousness

by Julian Keenan with Gordon G. Gallup Jr & Dean Falk

Harper Collins: 2003. 304 pp.

US\$24.95, Can\$38.95

Marc Hauser

In the film *Memento*, the main character, Leonard Shelby, suffers an injury to the head that creates a nightmarish problem. Although his experiences before the injury are clear, everything that has happened after it is unfamiliar, even events that occurred just moments before. Because the mind's recording device is broken, things are experienced but never recalled. Shelby's autobiography is frozen in the past.

It is hard for us to imagine what it would be like to lack this kind of memory. It would certainly rob us of one component of our sense of self, of what it was like to have experienced different things in the world. We would not be able to judge whether or not we like something; we would have to resample each event on the moment. Even if we resorted to writing down each experience, as Shelby tried to do, we would be constrained by the relatively impoverished connection between subjective experience and language.

Understanding how this sense of self is constructed and represented in the brain is the focus of *The Face in the Mirror*, a book written by three scientists who have all made significant contributions to their respective disciplines. Cognitive neuroscientist Julian Keenan is known for his neuroimaging studies of self-recognition; the comparative psychologist Gordon Gallup invented the mirror test, which has been used to explore self-recognition in animals and human infants; and the biological anthropologist Dean Falk has documented several important historical patterns of human brain evolution. Although they are all co-authors of the book, it is written in Keenan's first-person voice.

According to Keenan, the goal of the book is "to tell the story of self-awareness, why it exists, where it came from, and where in the brain we might locate this astonishing ability. The findings and conclusions we have reached are quite amazing and contrary to what many neuroscientists believe." This is a highly ambitious research programme. How well does he fare?

The book begins by laying the relevant groundwork, including the history of some of the ideas that drive research on the sense of self, in terms of both self-recognition and self-awareness. Although Keenan stays clear



A familiar face: our sense of self reflects our capacity for self-recognition and self-awareness.

of many of the interesting philosophical issues concerning subjective experience, and offers no insights into the selective pressures on or adaptive significance of our sense of self, the issues that he raises are nonetheless fascinating.

For example, when an individual recognizes herself in the mirror, what does this tell us, if anything, about what she is thinking? The answer to this question is important because it has implications for understanding the nature of animal and human minds. If, as Gallup proposed more than 20 years ago, mirror self-recognition provides a litmus test for self-awareness, then we can use it to explore which animals, and which humans, have this kind of a conscious mental life. If not, then other tests are needed to assess whether non-linguistic or pre-linguistic organisms have the capacity for self-awareness. Answering this question should also help us to refine our interpretation of neurophysiological and neuropsychological data from both humans and other animals.

Here, in a nutshell, is the reported evidence. First, not all animals show mirror self-recognition. This capacity seems to be restricted to the great apes and humans over the age of 18–24 months. Keenan claims that animals with self-recognition also have the capacity to attribute mental states to others — they have a theory of mind. More specifically, at about the time that children acquire mirror self-recognition, they also acquire a rudimentary sense of self-awareness. Consequently, so Keenan argues, mirror self-recognition

provides a litmus test for a theory of mind.

Second, patients with a loss of self-awareness often have deficits in mirror recognition. In some of these patients, especially those with anosognosia, which involves paralysis to one side of the body accompanied by complete denial that the deficit exists, the damage is typically in the right hemisphere. Third, neuroimaging and psychophysical studies indicate that the right hemisphere has a dominant role in self-recognition. So Keenan suggests that the capacity for self-recognition is deeply connected to the capacity for self-awareness; that this connection evolved before the emergence of our species; and that our sense of self resides in the right hemisphere.

This is an intriguing hypothesis, with some supporting evidence, but there are three problems. To begin with, there are some situations in which humans fail to show mirror self-recognition, but still maintain a rich sense of who they are and what they believe and feel. Some patients with prosopagnosia — a selective face-recognition deficit — fail to recognize their own mirror image or a picture of their own face, and this realization strikes at the core of their sense of self. So there is not a necessary relationship between self-recognition and self-awareness. Similarly, the evidence that Keenan cites supporting a relationship between the mirror test and a theory of mind in apes has been dismissed as inconclusive by others, including Daniel Povinelli, who carried out the original research on the idea. Oddly, Keenan does not mention

the much stronger evidence that has been compiled in the past three years by Brian Hare, Michael Tomasello and Josep Call, but even this work says nothing about the content of a chimpanzee's beliefs.

A second problem is that research on the developing theory of mind goes well beyond the literature reviewed by Keenan, and raises significant problems for his theory. Specifically, it is now clear that for many cognitive capacities, including the theory of mind, children have implicit knowledge long before it is explicit; using an explicit measure such as the mirror test might cause us to miss an earlier capacity. Furthermore, part of the challenge for the child in acquiring an explicit theory of mind is that it requires considerable executive control, something that young children lack because of the relative immaturity of the frontal lobes. This means that a child's inability to understand what others believe is not necessarily a reflection of the challenges it faces in attributing similar or different beliefs to others; rather, the difficulty may stem from the challenge associated with inhibiting personal beliefs in order to make accurate judgements about others.

Third, even if our sense of self is located in the right hemisphere — and there is chauvinism against this side of the brain from neuroscientists who think that the left hemisphere does all the heavy intellectual lifting — this does not help our understanding of how the brain generates a feeling of personal experience, of guilt, awe, shame or despair. It wouldn't help us to understand why Shelby

is distraught, or whether medical technology might someday reverse these cases of brain damage.

In the end, however, Keenan and his co-authors have assembled a rich set of evidence that will contribute to what is certainly one of the most interesting topics in the sciences of the mind: our sense of self. At least, that's what I think, I think. ■

Marc Hauser is at the Primate Cognitive Neuroscience Laboratory and the Department of Psychology, Harvard University, Cambridge, Massachusetts 02138, USA.

A flexible theory of evolution

Developmental Plasticity and Evolution

by Mary Jane West-Eberhard
Oxford University Press: 2003. 720 pp. \$100, £79.50 (hbk); \$49.95, £37.50 (pbk)

Gerdien de Jong and Ross H. Crozier

It is an ancient truism that what evolves is the developmental system, from which it follows that genetics, development and evolution are interwoven. Genetics and evolution were integrated long ago in the synthetic theory of evolution; development has only lately rejoined the evolutionary fold. The field of evolutionary developmental biology thus created is mostly concerned with patterns of evolution, comparing the genetic basis of evolutionary changes in development, rather than with the dynamics of evolutionary changes. Its sister, developmental evolutionary biology, studies how selection works on development to produce

the adaptive phenotype, necessarily becoming linked to evolutionary ecology. This is a field worth pursuing, and Mary Jane West-Eberhard's *Developmental Plasticity and Evolution* certainly belongs within it, but the question is whether the book does the field good service.

For West-Eberhard the phenotype is central. But this is not the bleak 'genotype + environment = phenotype' taught to first-year students, but rather a vibrant, living, changing phenotypic whole, far from dreary genetic determinism. The phenotype is developmentally plastic, changing in many ways in response to many environmental challenges. To be alive is to be developmentally plastic. West-Eberhard envisages a synthetic theory of evolution and development in which environmentally induced phenotypic change gives rise to adaptive evolution as readily as, or even more readily than, mutationally induced phenotypic change.

The main evolutionary process, in West-Eberhard's universe, involves environmental change, phenotypic accommodation and genetic accommodation. An environmental change elicits a developmentally plastic response, and phenotypic accommodation — the immediate adjustment to a change resulting from the multidimensional adaptive flexibility of the phenotype — ameliorates its harm to individuals. New phenotypes resulting from this developmental plasticity are selected. A change in allele frequency — genetic accommodation — improves and incorporates the change. In this way the environment becomes a crucial participant in the generation and selection of adaptive design.

In West-Eberhard's view, this sequence of developmental plasticity, phenotypic accommodation and genetic accommodation is the mechanism responsible for (nearly) all

New in paperback

Science, Money, and Politics: Political Triumph and Ethical Erosion

by Daniel S. Greenberg

University of Chicago Press, \$20, £14

"...the value of this book [is] as a unique and revealing perspective on the way that the science-funding process actually works in Washington. The picture it paints is not a flattering one. But — unlike many of those he writes about — Greenberg is not out to make friends in high places." David Dickson, *Nature* **413**, 355–356 (2001).

The Future of Life

by Edward O. Wilson

Abacus, £8.99

"[Wilson] accurately and passionately tells the story of the disappearance of many of the only living beings we know of in the Universe — key components of humanity's natural capital." Paul R. Ehrlich, *Nature* **417**, 21–22 (2002).

Secret Agents: The Menace of Emerging Infections

by Madeline Drexler

Penguin, \$15.00



Spot the difference: the water flea *Daphnia* readily changes shape in response to its environment.

CHRISTIAN LAFOURCH/SCIENCE PHOTO LIBRARY

evolutionary novelty, adaptive radiation, speciation and macroevolution. Evolution proceeds through adaptive developmental phenotypic plasticity.

Mainstream evolutionary biology is riddled with genetic determinism, blocking a biological view of the organic phenotype, West-Eberhard suggests. However, many of the biological studies she adduces in her support belong to the mainstream. Nobody objects to “a unified Darwinian theory that relates developmental plasticity to genetic change”, but it is a matter of proportion. West-Eberhard sees too much unwarranted emphasis on genes; many will see an unwarranted role for developmental plasticity in her argument, as the description of phenotypic plasticity itself is too general and vague to get to grips with.

West-Eberhard's conviction of the primacy of the environment as the inducer of new phenotypic variation runs through the book, making her ask for a coherent evolutionary theory that uncompromisingly includes the environment alongside the genome in all aspects of evolutionary thought. She accepts that genetic change accompanies evolution, but only as genetic accommodation follows environmental induction; evolution as genetic change “is left hanging by a tenuous thread”. In this view, genes are followers, not leaders.

How plausible is all this? Not very. No convincing evidence is presented for adaptive phenotypic accommodation to a new environment. Genetic accommodation is just a classical adaptive change in gene frequency. Developmental plasticity exists and is important in nature, but for it to be the dominant evolutionary factor, one has to show that developmental plasticity is predominantly adaptive and precedes genetic adaptation.

West-Eberhard refers to much good biology, but fails on the major point: developmental plasticity as the initiating factor of adaptive novelty preceding genetic change. The evidence shows that much developmental plasticity exists and has a genetic basis, no more and no less. West-Eberhard concedes that direct examples of adaptive environmental induction are lacking, but she labels many cases as indirect evidence reflecting the process. Whether the environment is the main player in eliciting adaptive developmental plasticity, and thereby in all other evolutionary processes, remains a question of faith. No crucial laboratory experiment is suggested that would test whether environmental induction leads to adaptive evolution. Actually, some tests for phenotypic accommodation (as ‘beneficial acclimation’) have been done by Raymond Huey's group at the University of Washington, and were negative.

In one of the first studies of phenotypic plasticity, Richard Woltereck defined the

reaction norm as the range of the phenotypes that an individual could exhibit over all environments. The original insight was: “*Genotypus = Reaktionsnorm*”, the genotype being the information for developmental plasticity. Woltereck transplanted *Daphnia* from Denmark to Italy to investigate whether the environment modified the reaction norm: it didn't. Later, dissatisfied with ‘materialism’ but impressed with phenotypic plasticity, Woltereck wrote two books, unfortunately incomprehensible, expressing a holistic view. In reading West-Eberhard's *Developmental Plasticity and Evolution*, one also often struggles with the verbal arguments: what does this mean, and how, precisely, would that work? West-Eberhard asserts a vision but presents little analysis. A major new synthesis and research programme this book is not.

Gerdien de Jong is in the Department of Evolutionary Population Biology, Utrecht University, Padualaan 8, NL-3584 CH Utrecht, the Netherlands; Ross H. Crozier is in the Department of Evolutionary Genetics, School of Tropical Biology, James Cook University, Townsville, Queensland 4811, Australia.

..... The finger of Galileo

The Oxford Companion to the History of Modern Science

Edited by J. L. Heilbron

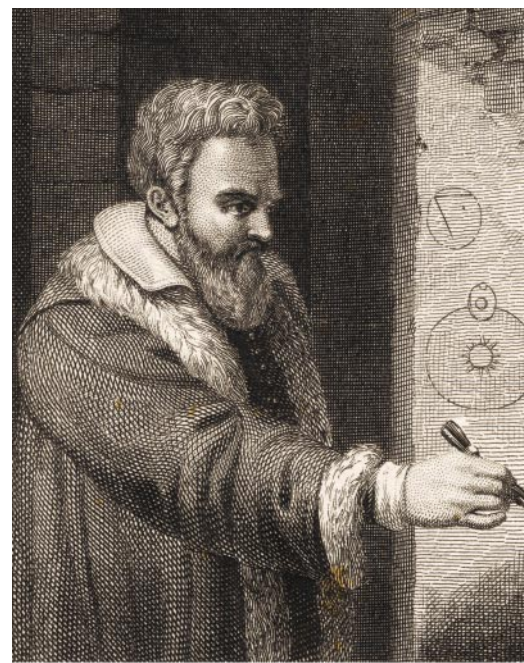
Oxford University Press: 2003. 960 pp.

\$110, £40

Ryan J. Huxtable

The value of a person's life is not to be judged by the length of the obituary. However, more than 100 biographies of scientists in *The Oxford Companion to the History of Modern Science* are all approximately one page in length, regardless of the achievement of the scientist or the amount of information available about the person. This modified egalitarianism is illuminating not only of the subject at issue, but of current attitudes towards personal achievement and great men or women. Linus Pauling, who uniquely won two individual Nobel prizes, made enormous contributions to several areas, including insights into the nature of the chemical bond, and lived to a ripe age of 93, still working on controversial issues. Rosalind Franklin made an important, but single, contribution to the elucidation of the structure of DNA, and died at the early age of 37. These two scientists get the same consideration as Galileo.

By almost any measure, modern science begins with Galileo Galilei (1564–1642). He made enormous theoretical and practical contributions to areas as disparate as astronomy and the measurement of time. He tore down the dusty shed of aristotelian physics



Galileo Galilei: making plans for the future.

that had blocked the daylight for so long. He was also one of the first rational investigators to fall foul of government and orthodoxy. His entry in this book describes him as the father of modern science, and the preface to the book begins with a quotation from him.

Conspicuously absent is an entry on Giordano Bruno, who was burnt at the stake for his iconoclastic but rational views of astronomy a few years before Galileo faced the Inquisition. Who can doubt that Galileo's recanting of his ‘heresies’ of heliocentrism was fuelled by the fire that consumed Bruno? Bruno receives three mentions *en passant*, on the influences on him, and on his views of an infinite Universe and stellar distances. Bruno was a hermeticist. Hermeticism, which does get a deserved entry in this book, was a part rational, part mystical system of thought, significant in the development of modern science. However, Galileo was surely the first modern scientist.

The editor-in-chief, J. L. Heilbron, is a witty and erudite contributor. His entry on ‘ether’ begins: “A possibly nonexistent entity”, a description that in itself conjures up various philosophical issues. Are Plato's ‘ideal horses’ entities? Heilbron is aided by 5 editors, 7 consultants and 217 contributors, in producing almost a thousand pages of double-column text.

The thematic listing at the front of the volume of main headings and subheadings was more useful than the index. Thus, an index entry of page 357 for ‘civil rights’ leaves the reader scanning two densely printed columns in search of the elusive reference.

Some of the entries, such as that on ‘progress’, seem overly general. Others seem non-intuitive, such as ‘tacit knowledge’, ‘shift

BETTMANN/CORBIS

Science in culture

Seeing sense

Annie Cattrell's sculptures of the five senses are on display at the Royal Institution in London.

Martin Kemp

The mission of the Royal Institution, founded in 1799, focuses on the aspiration to "diffuse knowledge, and, through philosophical lectures and experiments, apply science for the common purposes of life". Visual demonstration has always been central to this aim, not least through the popular discourses delivered by Michael Faraday in the nineteenth century. Faraday was apprenticed to a book-binder before rising to become the greatest hands-on scientist of his generation.

Among the visual wonders that Faraday demonstrated at the Royal Institution were the first exhibited 'photogenic drawings' (early photographs) by William Henry Fox Talbot. Their rushed display in January 1839 was triggered by the startling French announcement at the Academy of Sciences two weeks earlier of Louis-Jacques-Mandé Daguerre's 'invention' of what came to be called photography.

Fittingly, the exhibition "From Within" by Annie Cattrell, the Royal Institution's artist in residence last year, includes photograms (direct exposures) in the manner of Talbot. She has made images of a sectioned human skull, created by exposing the skull and its cranial cap directly over photographic paper and flooding its interior with light from a hand-held torch. The negative reversal inherent in these photograms eerily maps the contours and orifices of the cranium against a black substratum, and seems to reveal its cavernous interior as a radiant source of mental illumination. In the spirit of Faraday, Cattrell has also prepared a video of magnetized iron filings, ingenious visualizations in cut paper of frictional forces, and a small installation of images of water placed between the faces of cut diamonds and subjected to extreme pressures.

The brain itself is the subject of Cattrell's most sustained exploration of how abstract visualizations in science can be turned into tangible reality. Her set of cubic sculptures *The Five Senses* is the culmination of three years of intense research. Two of the sculptures were finished in time for the "Head On" exhibition at the Science Museum in London last year, and now all five are complete. They rework a long-standing iconographical theme, which proved particularly popular in prints from the Renaissance onwards.



In *The Five Senses*, Annie Cattrell explores the physical underpinnings of consciousness.

Among the texts that Cattrell studied was *The Human Brain* by Susan Greenfield, the Royal Institution's current director. Cattrell also discussed the work and collaborated with various brain scientists, including Steve Smith and Morten Kringelbach of the University of Oxford, and Mark Lythgoe of the Institute of Child Health in London, who granted her access to brain activity data generated by functional magnetic resonance imaging. The technique of rapid prototyping, courtesy of Californian company 3D Systems, translated the data into three-dimensional form.

Cattrell is seeking to grasp the "physicality of consciousness" by exploring the "delicate dialogue between the exterior world and our individual blueprint". She models this dialogue by casting in resin the morphological patterns of brain activity that correspond to the stimulation of each of the five senses. Neural activity is transformed into glistening apparitions that float in the cranial cavity like a kind of mental plasma. The refractive and reflective crystalline cubes, within which the skull is by implication inscribed, optically slice the golden configurations into shifting interplays of plans and elevations as the spectator moves past them.

In imaging the brain by casting and modelling, Cattrell stands in a long line going back to Leonardo da Vinci, who cast the ventricles of an ox brain, believing that the fluid in the ventricles was the medium within which the mental faculties operated. Of particular fascination to Cattrell are the almost unbelievably refined creations of the great wax modellers of the eighteenth and nineteenth centuries, including wax brains created by Joseph Towne that are in the Gordon Museum at Guy's Hospital, London. But whereas Towne's demonstrations can be characterized as pedagogy charged with beauty, Cattrell is in no sense working as an illustrator. Rather, as an artist she imaginatively translates the technical data, in all its awesome detail, into perceptible and beautiful forms that do full justice to the scientists' own excitement in creative visualization.

Martin Kemp is professor of the history of art at the University of Oxford and co-director of Wallace Kemp/Artakt.

Annie Cattrell's exhibition "From Within" can be seen at the Faraday Museum of the Royal Institution, London, until mid-September.

of hegemony', 'diffusion in the east' (which has nothing to do with gases), or 'brain drains and paperclip operations'. 'Standard model' discusses GUTs and TOEs but not body parts (TOEs being 'theories of everything' and GUTs referring to 'grand unified theories'). The value of the thematic listing is shown by finding 'tacit knowledge' as a sub-heading of 'Epistemology and methodology', which, in turn, is an entry under 'The body of scientific knowledge'.

This volume is the culmination of much scholarship and enormous effort (one rare error is a reference to the "noble" prize in the preface). The result is delightful to browse, but it is difficult to see how the book could be used systematically. It is of no help, for example, in tracing the history of anaesthesia. Unintentional insight into the planned use of the book is perhaps given by repeated phrases such as "depicts for a general audience", indicating an emphasis more on seeing and

hearing than on reading. Indeed, I cannot escape a feeling that the time for print publication of such texts is passing. Electronic publication would provide easier searching and updating, and could more easily accommodate changing fashions. In short, this is one of those useful books for which it may be hard to find a use. ■

Ryan J. Huxtable is professor emeritus in the Department of Pharmacology, University of Arizona, Tucson, Arizona 85724, USA.

Classified information

Summarize yourself in the form of a title of a paper in Nature.

Quaternary mammal from Argentina sheds light on dinosaur evolution.

Who has been the most important mentor in your career?

José Bonaparte, a well-known Argentine vertebrate palaeontologist and my PhD adviser. From him I learned how to work hard, to persevere, and to believe that no obstacle is insurmountable.

Whose graduate student would you most like to have been (historical impossibility notwithstanding)?

Leonardo da Vinci, who not only had a voracious appetite for knowledge, but integrated science and art like nobody else.

What book has been most influential in your scientific career?

E. O. Wiley's *Phylogenetics: The Theory and Practice of Phylogenetic Systematics* (J. Wiley & Sons, 1981), a book that made me realize the importance of framing my research (and my way of interpreting the history of life) within cladistic methodology.

What was the worst/most memorable comment you ever received from a referee?

A reviewer of one of my books said that he wanted to throw it away after reading the first few pages.

You have the audience in your hands, but some smart-alec asks you the killer question you have no idea how to answer. What's your response?

Something straight: "Sorry, you've got me — I have no clue."

What book currently resides on your bedside table?

The Riddle of the Compass by Amir Aczel, and Michael White's *Leonardo: the First Scientist*.

What music heads the playlist in your car or laboratory?

All sorts of things, from Handel to Joan Manuel Serrat to Cat Stevens. It depends on the mood of the day.

What do you most dislike about having research published?

Finding mistakes that I could have corrected.

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

The Renaissance painter Hieronymus Bosch, the Chilean poet Pablo Neruda, and Marco

Polo. In many ways, all of them thought differently from their contemporaries.

Where and when would you most like to have lived or worked?

In North America or Europe around the turn of the twentieth century, when much more uncharted territory and funds were available for palaeontological exploration and discovery.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

I would remain quiet and listen. It might give me another opportunity to learn how to improve my work and how to communicate it better.

The Internet is the bane of scientists' lives because...

...e-mail takes up a large amount of time and still leaves you with the sour feeling of not being able to handle all your correspondence. I fear that the Internet has made available a large amount of misinformation, through lists, amateur pages and publications that are not peer-reviewed.

What do you do to relax?

I like to look at paintings. What I like most is to find a quiet bench at an art museum where I can relax in the company of beautiful canvases. I also like to take care of my garden and to hang out at the beach at sunset. Doing fieldwork also relaxes me, although it brings other kinds of stress.

What would you have become, if not a scientist?

In retrospect — I did not entertain this when I went to university — an art historian. It is the historical aspect that interests me most, whether the object is man-made or of natural design.

What's the one thing about science that you wish the public understood better?

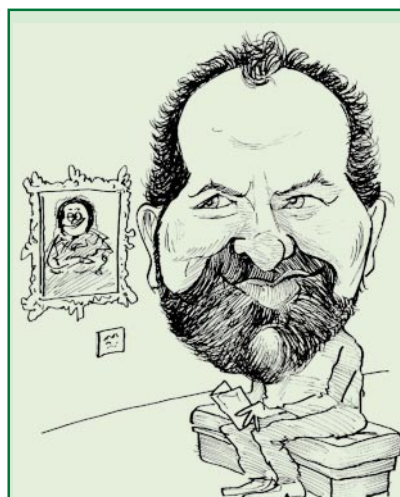
The importance of systematics. How our understanding of the evolutionary history of organisms should underpin all other investigations about them.

What's your motto?

It would be something along the lines of Henri Matisse's "What I had to do, I did the best I could."

The job of captain on the Starship Enterprise in Star Trek has become vacant. Nominate any real person, living or dead, for the post.

Roy Chapman Andrews, who led the central Asiatic expeditions to the Gobi Desert in the



Luis Chiappe

Luis Chiappe is a curator and chair of the department of vertebrate palaeontology at the Natural History Museum of Los Angeles County. When he is not doing research on dinosaur evolution, he enjoys art, gardening, the outdoors, fishing and living by the sea.

1920s. Someone who had not just a great spirit of adventure and discovery, but who also who knew how to make that excitement contagious.

Do you have a burning ambition to do or learn something of no practical or immediate value? If so, what?

I would like to sail to a remote island, a trip that would take days of navigation to reach my destination, and would give me a lot of time to reflect on my work.

Harry Potter or Lord of the Rings?

I haven't read Harry Potter (nor have I seen the movies) but I have always treasured J. R. R. Tolkien's work — *The Lord of the Rings*, *The Hobbit* and *The Silmarillion* are among my favourite books. As someone used to describing objects (fossil specimens), I have always admired Tolkien's outstanding ability to describe the most fascinating landscapes.

What's just around the corner?

Affordable three-dimensional scanners that can scan objects of all sizes. Many museums are already digitizing their collections but in a two-dimensional, photographic format. As three-dimensional scanners become affordable and museums begin to digitize their specimens in three-dimensional files that can be downloaded and replicated by prototyping systems, the nature of how research is conducted in specimen-based institutions will change drastically. ■

Good to CU

Yapeng Gu and Wesley I. Sundquist

A powerful arm of the cellular defence against microbial invaders has been characterized. APOBEC3G, a protein that can fight off HIV, works by introducing 'typographical errors' during viral replication.

HIV encodes a protein that allows the virus to multiply in otherwise resistant human cells¹⁻³. This protein, Vif — for 'viral infectivity factor' — works by overcoming a cellular protein, APOBEC3G, whose task is to inhibit the replication of HIV and other retroviruses⁴. As they describe in *Cell*⁵ and on pages 94 and 99 of this issue^{6,7}, three groups have now discovered how APOBEC3G hampers viral replication. The authors find that, as the retroviral RNA genome is copied into DNA in host cells, this protein changes 'C's to 'U's in the DNA. APOBEC3G therefore represents a remarkable innate cellular defence mechanism that drastically alters the chemical composition and coding capacity of replicating retroviruses.

Retroviruses, which include HIV and murine leukaemia virus (MLV), have genomes that are made up of single-stranded RNA. For these viruses to replicate in a host cell, a double-stranded DNA copy of the RNA genome must be made, by the process of reverse transcription. This happens in two steps. First, a DNA strand that has a sequence complementary to the RNA is produced (this is called 'minus-strand' DNA). Then the RNA is removed and a second (plus-strand) DNA molecule, complementary to the first (and therefore matching the sequence of the

original RNA), is synthesized. The resulting double-stranded DNA is incorporated into the host-cell genome, and, when switched on, is used to generate the proteins and RNA needed to form new virus particles (virions).

It has been known for more than a decade that the Vif protein is required for HIV to replicate in some human cell types (termed 'non-permissive' cells), but not others ('permissive' cells)¹⁻³. Non-permissive cells express APOBEC3G (also known as CEM15), and permissive cells become non-permissive upon expression of this protein⁴. These and other data indicate that APOBEC3G blocks viral replication, and that Vif works by overcoming this block. However, the precise nature of the replication block has not been clear — and it is unusual in that it is manifest after the virus has entered a new target cell. So, Vif-deficient viruses produced by non-permissive cells are released at normal levels, but are somehow impaired in their ability to form stable, productive replication intermediates when they enter new target cells^{3,8}.

Clues to how this might happen have come from several observations. First, APOBEC3G can be packaged into HIV-1 virions⁴. Second, APOBEC3G is related to a family of enzymes that catalyse the deamination of cytosine bases (C's) in DNA and RNA, thereby producing uracils (U's). Finally, APOBEC3G can mutate the DNA of

the bacterium *Escherichia coli*⁹. Thus it was suggested that APOBEC3G might inhibit HIV replication by altering the coding capacity of viral RNA or DNA⁴.

Harris *et al.*⁵, Mangeat *et al.*⁶ and Zhang *et al.*⁷ have now shown that APOBEC3G can indeed alter the sequences of the HIV and other retroviral genomes, producing an exceptionally high frequency of guanine-to-adenosine (G-to-A) substitutions in the plus (protein-coding) strand of the viral DNA. As shown in detail in Fig. 1, this observation implies that APOBEC3G changes C's to U's in the DNA minus strand during replication (because, when the plus strand is made using the minus strand as a template, G's are inserted opposite C's, but A's opposite U's). Importantly, G-to-A hypermutation is suppressed by the presence of Vif^{5-7,10}, and APOBEC3G has the required deaminase activity^{5,7} and can act on single-stranded DNA substrates⁵. Deamination by APOBEC3G is quite promiscuous, but exhibits some sequence selectivity, with a preference for C/T-C-C sequences⁵⁻⁷.

Conversion of C's to U's in the genomic DNA minus strand apparently blocks viral replication in several ways (Fig. 1). First, the presence of uracils can target the minus strand for destruction. Uracil is not normally found in DNA and can therefore be excised by host DNA-repair enzymes. This could

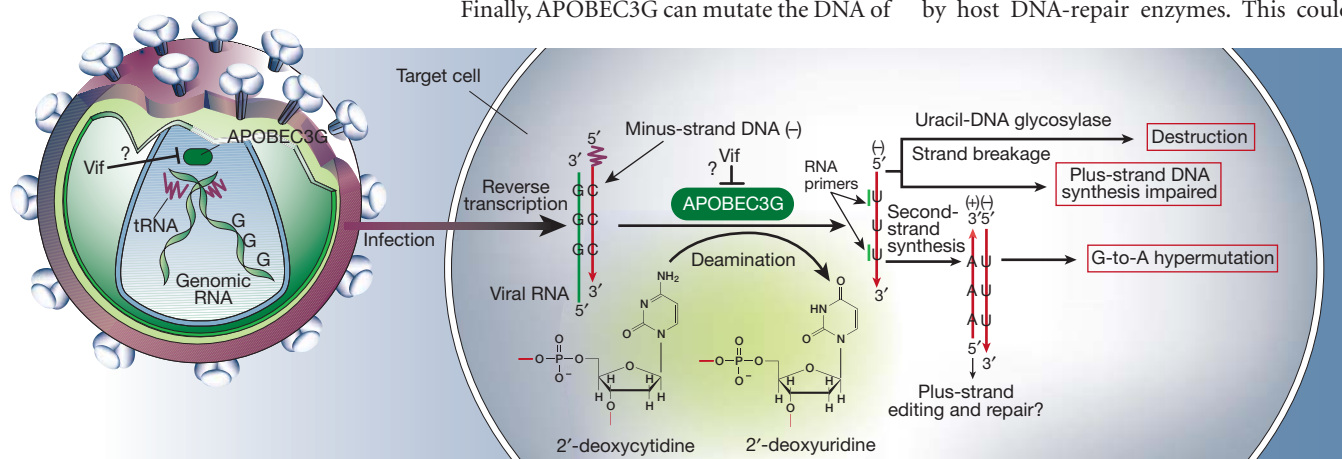


Figure 1 Editing HIV: how the cellular APOBEC3G protein might inhibit HIV replication^{5-7,10}. From left, APOBEC3G has been packaged, together with the viral genome, in a viral particle. Then, when the retrovirus enters a new cell, the viral reverse transcriptase copies the genomic RNA into a double-stranded DNA. Replication begins at a tRNA primer (purple), and creates first a DNA minus (–) strand that is a complementary copy of the RNA, and then a DNA plus (+) strand that is a complementary copy of the minus strand. However, APOBEC3G deaminates 2'-deoxycytidines (C) in

the minus strand, producing 2'-deoxyuridines (U). Plus-strand synthesis then converts these C-to-U changes into G-to-A mutations, because C pairs with G, whereas U pairs with A. C-to-U conversion can diminish HIV-1 replication in several ways: deglycosylation of uracil can lead to strand breakage and destruction; the presence of U's can reduce plus-strand synthesis¹¹; and G-to-A hypermutations in the viral genome can impair all subsequent viral functions. The viral Vif protein blocks the effects of APOBEC3G through an as yet undefined mechanism.

eventually lead to strand breakage and destruction — explaining the apparent instability of viral reverse transcripts. Second, uracils in the minus strand can impair the initiation of plus-strand synthesis during HIV-1 reverse transcription¹¹. Finally, even when complete, double-stranded DNA copies of the viral genome are made, they contain many G-to-A mutations, which result in amino-acid changes and aberrant 'stop' signals in the encoded proteins, and generally reduce viral fitness at every subsequent stage in replication⁶. The biased G-to-A mutation profile might also explain the overall A-richness of the HIV-1 genome, as well as the non-lethal hypermutation sometimes observed during viral replication^{5–7,10}.

As with all other immune responses, APOBEC3G must successfully distinguish between 'self' and 'non-self'. So it is remarkable that, as two of the groups show, the protein can inhibit the infectivity of retroviruses that differ markedly in sequence, including MLV^{5,6}, simian immunodeficiency virus, equine infectious anaemia virus (EIAV), and even engineered retroviruses — derived from HIV-1 — that contain little of the original genome⁶. Presumably, these genomes are specifically targeted because APOBEC3G is incorporated into virus particles, and/or because only virus-specific intermediates are selected for deamination. It will therefore be important to learn how APOBEC3G is packaged into virions, and how it selects its substrates. Intriguingly, another member of this enzyme family, AID (activation-induced cytidine deaminase), which is involved in antibody-gene diversification, also targets single-stranded DNA. It requires transcription of the targeted gene into RNA^{12,13}, and the presence of an RNA-degrading enzyme¹³. By analogy, APOBEC3G might target single-stranded minus DNA after reverse transcription and removal of the RNA template by the RNA-degrading activity of the reverse transcriptase enzyme. Alternatively, substrate targeting could be mediated by a cofactor, as is the case for another family member¹⁴.

Another issue, which could have therapeutic implications, is how Vif overcomes the antiviral activity of APOBEC3G. Possibilities include downregulation of APOBEC3G protein levels, exclusion of the protein from progeny virions, and inactivation of incorporated APOBEC3G molecules⁴. Moreover, do viruses such as MLV and EIAV, which do not have the Vif gene, avoid the effects of APOBEC3G by expressing other proteins with Vif-like activities? Or do they simply replicate only in host cells that lack APOBEC3G?

It is becoming increasingly apparent that cells have evolved a remarkable number of mechanisms to defend themselves from microbial invaders, and that microbes have discovered clever ways to circumvent those defences. Retroviruses are particularly good

tools with which to study innate cellular immune systems, because the battle between cells and retroviruses is an ancient and intense one — as exemplified by the fact that retroviral elements make up an astonishing 8% of the human genome¹⁵. Other cellular defences against retroviruses include ZAP, a protein that targets viral messenger RNAs for destruction¹⁶, and Fv-1 (and related cellular proteins), which inhibits early stages of retroviral replication^{17,18}. The molecular description of APOBEC3G activity^{5–7,10} has revealed yet another arm of the innate immune system, in which cells actually edit the genomes of invading retroviruses. ■

Yapeng Gu and Wesley I. Sundquist are in the Department of Biochemistry, University of Utah, Salt Lake City, Utah 84132, USA.
e-mail: wes@biochem.utah.edu

1. Fisher, A. G. *et al.* *Science* **237**, 888–893 (1987).
2. Strebel, K. *et al.* *Nature* **328**, 728–730 (1987).

3. Gabuzda, D. H. *et al.* *J. Virol.* **66**, 6489–6495 (1992).
4. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. *Nature* **418**, 646–650 (2002).
5. Harris, R. S. *et al.* *Cell* **113**, 803–809 (2003).
6. Mangeat, B. *et al.* *Nature* **424**, 99–102 (2003).
7. Zhang, H. *et al.* *Nature* **424**, 94–98 (2003).
8. von Schwedler, U., Song, J., Aiken, C. & Trono, D. *J. Virol.* **67**, 4945–4955 (1993).
9. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. *Mol. Cell* **10**, 1247–1253 (2002).
10. Lecossier, D., Bouchonnet, F., Clavel, F. & Hance, A. J. *Science* **300**, 1112 (2003).
11. Klarmann, G. J., Chen, X., North, T. W. & Preston, B. D. *J. Biol. Chem.* **278**, 7902–7909 (2003).
12. Chaudhuri, J. *et al.* *Nature* **422**, 726–730 (2003).
13. Bransteitter, R., Pham, P., Scharff, M. D. & Goodman, M. F. *Proc. Natl Acad. Sci. USA* **100**, 4102–4107 (2003).
14. Mehta, A., Banerjee, S. & Driscoll, D. M. *J. Biol. Chem.* **271**, 28294–28299 (1996).
15. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
16. Gao, G., Guo, X. & Goff, S. P. *Science* **297**, 1703–1706 (2002).
17. Best, S., Le Tissier, P., Towers, G. & Stoye, J. P. *Nature* **382**, 826–829 (1996).
18. Hatzioannou, T., Cowan, S., Goff, S. P., Bieniasz, P. D. & Towers, G. J. *EMBO J.* **22**, 385–394 (2003).

Planetary science

The history of air

H. J. Melosh

Giant impacts on Earth destroyed the envelope of gases surrounding the fledgling planet — so how has the modern-day planet regained its atmosphere? The answer, it seems, is that all was not lost.

The Earth was born in violence. Modern scenarios of its origin suggest that, as the Earth grew, matter arrived in progressively larger chunks. The final crescendo included the impact of a Mars-size proto-planet that added mass and energy to the nascent Earth and, incidentally, created the Moon. So great was the energy delivered in this impact that the proto-Earth probably melted completely, and silicate vapour formed a fiery (although short-lived) envelope around the planet. Amidst such hostility, it seems hardly possible that the fragile envelope of atmospheric gases could survive. Planetary scientists have taken it virtually for granted that the primordial atmosphere of the proto-Earth would have been stripped by such a stupendous impact, and have looked for mechanisms that might have regenerated the gaseous envelope after the tumult subsided. But Genda and Abe have taken a closer look at the problem and, writing in *Icarus*¹, they show that Earth's atmosphere is not quite as fragile as it once seemed.

The idea that impacts of kilometre-size comets or asteroids might strip off a small fraction of a terrestrial planet's atmosphere has been discussed for many years (for example, see ref. 2). In these relatively small events, the atmosphere in the vicinity of the impact is driven off by the high-speed vapour and debris that are ejected from the crater³. For planet-size impactors, however, the loss mechanism is different. Of course,

the atmosphere close to the impact site will be ejected with the other high-speed gases from the vaporized projectile. But, because the depth of the atmosphere is only a small fraction of the Earth's radius, this kind of direct stripping cannot remove the atmosphere that lies far from the impact site. Instead, a planetary-scale impact creates a strong shock wave in the Earth's mantle that propagates through its dense interior and emerges as a sudden jump in surface velocity at locations far from the impact site. If the velocity jump is big enough, a further shock wave forms that may accelerate the atmosphere to a high enough velocity for it to escape the Earth.

Chen and Ahrens⁴ were the first to analyse this global atmospheric ejection process. They assumed that the velocity of the motion generated at the surface reaches 8 km s⁻¹ and found that, at this velocity, most of the atmosphere is ejected. Genda and Abe¹ used a one-dimensional atmosphere model similar to that of Chen and Ahrens, but they benefited from more recent giant-impact modelling that shows that the actual surface velocities following a Mars-size impact are smaller than those assumed by Chen and Ahrens. According to such computations⁵, the maximum surface velocities only reach about 6 km s⁻¹ at the antipode of the impact and are smaller elsewhere. The difference in velocity between 6 and 8 km s⁻¹ may sound trivial, but it corresponds to a steep step in

the curve of ejection fraction versus velocity and so makes a big difference: Chen and Ahrens found that “almost all atmosphere” is lost at 8 km s^{-1} , but Genda and Abe find that less than 30% of the atmosphere is lost at 6 km s^{-1} .

Why is this result important? Although no one has yet found a convincing and general explanation for how the terrestrial planets acquired their atmospheres, the present abundances, particularly of Earth’s heavy noble gases (neon, argon, xenon and krypton), seem hard to reconcile with a nearly complete loss of atmosphere after the Earth was assembled⁶. The abundances of these gases and their isotopes in the present atmospheres of Earth, Mars and Venus (insofar as we know them) differ substantially from their abundances either in the Sun’s atmosphere or in meteorites. These differences have led atmospheric scientists to postulate a wide variety of mechanisms by which such gases may be acquired, partially lost and isotopically fractionated⁷. But the mechanical ejection of gases that was previously imagined for the giant impact is no help at all with this problem: in the impact scenario, the gases are ejected wholesale, without separation or fractionation.

If Earth’s primordial atmosphere had been completely lost in the impact, then a new inventory of gases would have had to have been acquired later. It is unclear where such a secondary atmosphere would come from. It seems nearly certain that it would have a very different composition from that of either Earth’s original atmosphere or the original atmospheres of the other planets. Total atmospheric loss thus adds a major wild card to the already highly uncertain mix of constraints on atmospheric evolution.

By showing that total atmospheric loss in a giant impact is unlikely, Genda and Abe¹ have permitted us to think what once seemed unthinkable: in spite of the violence of the Moon-forming giant impact, our primordial atmosphere may have survived all of the vicissitudes of Earth’s turbulent birth. Going even further, the authors argue that much of the Mars-size impactor’s atmosphere would also have survived, merging with Earth’s own gas envelope.

Much remains to be done before we can understand how Earth’s atmosphere evolved to its present state. Genda and Abe’s computation could be augmented by a better equation of state for atmospheric gases at the high temperatures reached in the shock propagating upwards from the surface. A better understanding of how the solid Earth would respond to the shock of a giant impact would inspire more confidence in the correctness of the surface boundary conditions. And we are still a long way from understanding the processes that affect the composition of the atmosphere. Nevertheless, Genda and Abe’s paper represents a

major step forward in sorting out the history of the air we breathe.

H. J. Melosh is at the Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA.

e-mail: jmelosh@lpl.arizona.edu

1. Genda, H. & Abe, Y. *Icarus* **164**, 149–162 (2003).

2. Ahrens, T. J. *Annu. Rev. Earth Planet. Sci.* **21**, 525–555 (1993).
3. Melosh, H. J. & Vickery, A. M. *Eos* **69**, 388 (1988).
4. Chen, G. Q. & Ahrens, T. J. *Phys. Earth Planet. Inter.* **100**, 21–26 (1997).
5. Cameron, A. G. W. *Lunar Planet. Sci. Conf. XXIII*, 199–200 (1992).
6. Pepin, R. O. *Icarus* **126**, 148–156 (1997).
7. Pepin, R. O. *Icarus* **92**, 2–79 (1991).

Evolutionary biology

Polygamy and parenting

Mark Pagel

In most animal groups, females put more effort into rearing children, and males compete for female attention. But what about seahorses and pipefish, in which males invest the most in offspring?

Modern men who think that they can attract women by being good with children may wish to read a study by Wilson and colleagues¹ that has just appeared in the journal *Evolution*. In seahorses and pipefish at least, females compete for males only when they — the females — have the time to, and not, it seems, according to the effort the males put into looking after offspring.

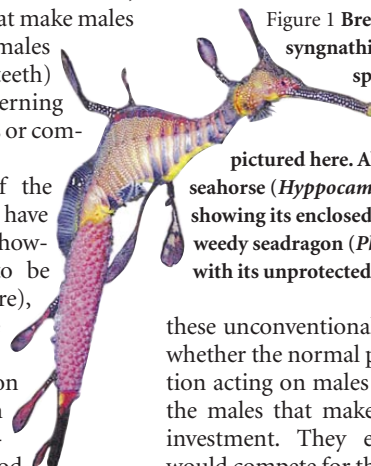
Evolutionary theory suggests that the relative amount of effort, or ‘investment’, that males and females put into rearing their offspring determines which sex competes for the other². In most animal species, females contribute substantially more to the offspring than males, and it is the males that compete for mates. This leads to a form of biological evolution called sexual selection, in which traits are favoured that make males good at competing with other males (such as big antlers or sharp teeth) or more attractive to discerning females (such as colourful tails or complex songs and displays).

Pipefish and seahorses of the family Syngnathidae (Fig. 1) have some rather different habits, however. Many species appear to be monogamous (a rarity in nature), and all male syngnathids have some form of specialized egg-brooding structure located on the abdomen or tail, into which females deposit their eggs during mating. The males’ brood pouches vary in complexity among species, from simple sticky patches to complex uterus-like structures with placenta-like features for nourishing the eggs: a mere transvestite seems unadventurous by comparison. By adopting these normally female traits and habits, males can be sure of their paternity of the fertilized eggs. But they achieve this by investing considerably more parental effort than females — and, it is assumed³, increasingly so as the brood pouch gets more complex.

Wilson and colleagues¹ took advantage of



Figure 1 Breaking the mould: syngnathid species. Males of these species put more effort than females into rearing offspring, as pictured here. Above, the pot-bellied seahorse (*Hippocampus abdominalis*), showing its enclosed brood pouch. Left, the weedy seadragon (*Phyllopteryx taeniolatus*), with its unprotected egg-compartments.



these unconventional animals to investigate whether the normal patterns of sexual selection acting on males get reversed when it is the males that make most of the parental investment. They expected that females would compete for the males (a pattern they call sex-role reversal) in the species with the most complex brood pouches. But they don’t. Instead, sex-role reversal in syngnathids tends to arise in those species with polygamous mating systems — those in which both males and females can have multiple mating partners. More tellingly, it is often the females of these species, and not the males, that attempt to have more than one partner, a mating system known as polyandry. So it appears that, in seahorses, natural selection for sexually exuberant female mating habits

R. KUTTER

drives sexual selection for female–female competition, independently of males' paternal investment. Selection for female polygamy also favours other traits that are normally associated with sexual selection in males of different animal groups: some female seahorses are vividly coloured.

Contrary to expectations, then, the assiduousness of syngnathid males does not predict sex-role reversal. But why? The answer is probably that about half of the species with complex brood pouches are monogamous, and that complicates any attempt at prediction, because females of monogamous species are generally just too busy to compete for other matings. Far from being the expression of undying mutual commitment and affection heralded by church and state, monogamy, if it even exists⁴, is a sort of evolutionary last resort: it arises only when both partners' full efforts are required to raise offspring successfully. Under these circumstances, both partners will be selected to evolve whatever adaptations will improve the offspring's survival. This may explain why the males of some seahorse species have evolved such elaborate brood pouches. It could also explain why female syngnathids are rarely sex-role-reversed in the monogamous species, even in those with the most complex brood pouches — these females are probably too encumbered helping to rear offspring to fight among themselves for males.

Is the picture that emerges from syngnathids — that sexual selection is determined not by relative amounts of parental effort, but rather by mating habits — likely to be general? The unusual propensity of seahorses and pipefish for monogamy may make their case exceptional. But it shows us that for relative degrees of parental investment to influence sexual selection, the sex that invests the least must have some flexibility in its mating habits. In support of this, some evidence from human groups suggests that as the strictures of monogamy are relaxed, men will take advantage of women who invest in children at high levels, by proportionately reducing their own contribution⁵. This, then, may become the incipient state of sexual selection for male–male competition (and likewise perhaps for female–female competition in the polygamous seahorses).

Finally, what are we to make of seahorse and pipefish males, especially those in the sex-role-reversed species: in what sense are they 'male'? Evolutionary biology prosaically defines maleness and femaleness by the size of one's gametes. Females make the large, well-provisioned and energetically costly gametes called eggs, and males make the tiny, unprovisioned sperm. Given that male syngnathids have found themselves in most other respects in the role of females, and vice versa, what prevents them from being not just sex-role-reversed but instead fully sex-reversed, with 'males' evolving egg-like

gametes and 'females' evolving sperm-like sex cells? Were females to make smaller eggs, they could produce more gametes in total to assist them in their efforts to mate with many males. Males in turn might be selected to increase the size of their sperm to ensure that the fertilized eggs were still well provisioned. The syngnathid sexual circle would then be complete, as girls became boys and boys became girls.

Has this ever happened? Whatever the answer, the inverted world of seahorses shows that simple gamete-based definitions of 'male' and 'female' can be misleading if taken to indicate a typical sex-role. Indeed, the genetic sex-determining architecture is intriguingly variable in animals⁶. Maybe

seahorses, with their occasional sex-role reversal and would-be-female males, show us one reason why.

Mark Pagel is in the School of Animal and Microbial Sciences, University of Reading, Reading RG6 6AJ, UK.

e-mail: m.pagel@rdg.ac.uk

1. Wilson, A. B., Ahnesjö, L., Vincent, A. C. J. & Meyer, A. *Evolution* **57**, 1374–1386 (2003).
2. Trivers, R. L. in *Sexual Selection and the Descent of Man* (ed. Campbell, B.) 136–179 (Heinemann, London, 1972).
3. Bergland, A., Rosenqvist, G. & Svensson, I. *Mar. Ecol. Prog. Ser.* **29**, 209–215 (1986).
4. Møller, A. P. in *The Encyclopedia of Evolution* (ed. Pagel, M.) 347–349 (Oxford Univ. Press, 2002).
5. Kaplan, H. S. & Lancaster, J. B. in *Offspring* (eds Wachter, K. et al.) 170–223 (Nat'l Academies Press, Washington DC, 2003).
6. Bull, J. J. in *The Evolution of Sex and its Consequences* (ed. Stearns, S. C.) 93–115 (Birkhäuser, Basel, 1987).

Condensed-matter physics

Really cool molecules

Paul S. Julienne

Ultracold molecules have been made by applying a changing magnetic field to a quantum gas of 'fermionic' atoms. This raises the prospect of creating novel superfluids and molecular Bose–Einstein condensates.

At the quantum level, any particle can also be considered to be a wave, its momentum corresponding to a wavelength. In ultracold atomic gases, at temperatures much lower than a millionth of a degree above absolute zero, the wavelength of the atoms becomes larger than the mean distance between them, giving rise to some remarkable quantum mechanical properties that are at the forefront of contemporary physics research¹. The properties of such cold quantum gases depend on whether the atoms are bosons or fermions. Bosonic atoms have integer values of the quantum number known as spin, and can form a Bose–Einstein condensate in which the trapped atoms occupy the single ground state of the system. Fermionic atoms, on the other hand, have half-integer spin values and each identical fermion must occupy a different quantum level. But if two fermions paired up to make a bosonic molecule, an exotic kind of superfluidity — flow without resistance — could be the result^{2–4}. On page 47 of this issue, Regal *et al.*⁵ describe an important step in this direction.

Their experiment began with a quantum gas of fermionic potassium atoms, a mixture of equal numbers of atoms with spin quantum numbers $-9/2$ and $-5/2$. The different spin states are necessary for the fermions to undergo collisions with each other with low collision energy. To create molecules from these atoms, Regal *et al.* took advantage of a special molecular state known as a Feshbach resonance, which may be thought of as a weakly bound pair of atoms. The energy of such a state can be tuned, using a magnetic

field, to lie close to that of two separated atoms. By ramping the magnetic field so that the energy of the resonance state moved from being above to being below the energy of two separated atoms, colliding pairs of atoms were induced to form a lower-energy molecule that is bound with respect to the separated atoms (Fig. 1). Regal *et al.* were able to convert up to half of the atoms to diatomic molecules.

Producing a molecular gas in this ultracold regime has proved to be a challenge. There have been several proposals to make molecules, and possibly a molecular Bose–Einstein condensate, by combining bosonic atoms in a Bose–Einstein condensate using a molecular resonance state^{6–10}. Experiments have come tantalizingly close. Wynar *et al.*¹¹ were the first to infer the formation of molecules nearly at rest, by observing the loss of condensed rubidium-87 atoms as they were photoassociated into molecules using a two-colour pulse of light. Donley *et al.*¹² have also given evidence for coherent atom–molecule interconversion^{13,14} in experiments with a rubidium-85 condensate that involved manipulating the energy of a Feshbach-resonance state using a sequence of magnetic-field pulses. However, neither experiment could detect any molecules directly: the imaging methods used for atoms do not work for molecules, because they have very different light-scattering properties.

Part of the beauty of the new work by Regal *et al.*⁵ is that they have conclusively demonstrated the presence of cold molecules. They exposed the molecules to a radio-frequency electromagnetic field of the right frequency to make them fall apart into single atoms with

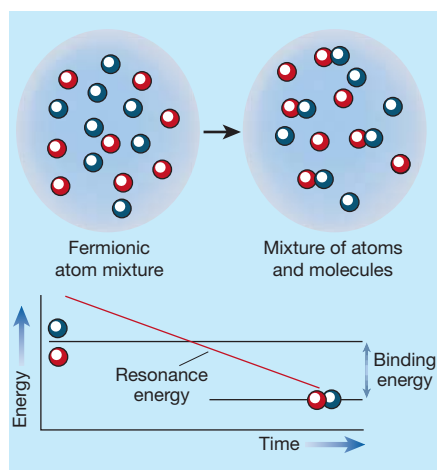


Figure 1 Making cold molecules. Regal *et al.*³ have succeeded in creating molecules in a mixture of cold gas atoms. The atoms exist in two different spin states (represented by red and blue dots) at a certain energy, but can be made to pair up through a Feshbach resonance. Varying the magnetic field applied to the system changes the energy of the resonance. When the resonance energy becomes the same as the energy level of the atomic mixture, colliding atoms can be converted to resonant-state molecules. As the resonance energy decreases further, the molecules finally arrive in a lower-energy state, lower than the atomic state by the amount of molecular binding energy.

spin quantum numbers $-9/2$ and $-7/2$. It is quite easy to detect atoms in different spin states separately, and as there were no $-7/2$ atoms present in the original mixture (only $-9/2$ and $-5/2$), counting the number of such atoms determines the number of molecules that had been present. By measuring the response of the molecules to different field frequencies, Regal *et al.* were also able to measure the extremely small binding energy of the molecules — of the order of only a nanoelectronvolt, and in excellent agreement with calculations based on the known properties of low-energy collisions of potassium atoms.

This experiment opens up several promising avenues for development. The molecules should have a temperature comparable to that of the atoms from which they were made — about 150 nanokelvins. A study of the momentum distribution and the coherence properties of such a molecular gas is clearly in order, to see whether it might actually be a molecular Bose–Einstein condensate. There are also challenges to theory, such as understanding why only half of the atoms were converted to molecules. Another major experimental goal is to demonstrate the existence of superfluidity in quantum degenerate fermionic gases, which might arise through the pairing of fermions in different spin states, in the same way that Cooper pairing of electrons is the basis of superconductivity. Theory even suggests that the very strong coupling associated with a Feshbach-resonance state should make

such an effect much easier to observe⁴.

A new era of research into ultracold molecular gases may be about to begin. There is much to do. For example, little is known about the collisions or chemical reactivity of ultracold molecules, which will determine the lifetime of such gases. Ultracold molecules also have potential uses in precise time and frequency measurements, in the search for an electron dipole moment¹⁵ and in quantum computing¹⁶.

Paul S. Julienne is in the Atomic Physics Division, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland 20899-8432, USA.

e-mail: paul.julienne@nist.gov

1. Nature Insight on Ultracold Matter *Nature* **416**, 205–246 (2002).

2. Houbiers, M. & Stoof, H. T. C. *Phys. Rev. A* **59**, 1556–1561 (1999).

3. Bruun, G., Castin, Y., Dum, R. & Burnett, K. *Eur. Phys. J. D* **7**, 433–439 (1999).

4. Holland, M., Kokkelmans, S. J. J. M. F., Chiofalo, M. L. & Walser, R. *Phys. Rev. Lett.* **87**, 120406 (2001).

5. Regal, C. A., Ticknor, C., Bohn, J. L. & Jin, D. S. *Nature* **424**, 47–50 (2003).

6. Julienne, P. S., Burnett, K., Band, Y. B. & Stwalley, W. C. *Phys. Rev. A* **58**, R797–R800 (1998).

7. Timmermans, E., Tommasini, P., Côté, R., Hussein, M. & Kerman, A. *Phys. Rev. Lett.* **83**, 2691–2694 (1999).

8. Heinzen, D. J., Wynar, R., Drummond, P. D. & Kheruntsyan, K. V. *Phys. Rev. Lett.* **84**, 5029–5032 (2000).

9. Mies, F. H., Tiesinga, E. & Julienne, P. S. *Phys. Rev. A* **61**, 022721 (2000).

10. Kokkelmans, S. J. J. M. F., Vissers, H. M. J. & Verhaar, B. J. *Phys. Rev. A* **63**, 031601 (2001).

11. Wynar, R. H., Freeland, R. S., Han, D. J., Ryu, C. & Heinzen, D. J. *Science* **287**, 1016–1019 (2000).

12. Donley, E. A., Claussen, N. R., Thompson, S. T. & Wieman, C. E. *Nature* **417**, 529–533 (2002).

13. Kokkelmans, S. J. J. M. F. & Holland, M. J. *Phys. Rev. Lett.* **89**, 180401 (2002).

14. Köhler, T., Gasenzer, T. & Burnett, K. *Phys. Rev. A* **67**, 013601 (2003).

15. Hudson, J. J., Sauer, B. E., Tarbutt, M. R. & Hinds, E. A. *Phys. Rev. Lett.* **89**, 023003 (2002).

16. DeMille, D. *Phys. Rev. Lett.* **88**, 067901 (2002).

Reproductive biology

Mammary messages

Elliott M. Blass

Identification of a pheromone that induces suckling in newborn rabbits sets a standard for studies on other mammals, and should prime investigations of the neurobiological basis of this behaviour.

One of the unique features of mammals is the parental nursing of young and suckling of milk from the mother¹. On page 68 of this issue², Schaal *et al.* describe a major step towards gaining a deeper understanding of the nursing–suckling relationship, and of pheromones. Pheromones elicit specific behavioural or physiological activity between members of the same species³, and their effectiveness does not depend on learning. In mammals, most research has centred on pheromonal influences on reproductive behaviour, and one such aspect has been tackled by Schaal and colleagues. They have identified a substance — 2-methylbut-2-enal (2MB2) — that is synthesized in the mammary gland of rabbit mothers and is released in milk to elicit the complex behavioural sequence that culminates in the attachment of a pup to the mother's nipple, and suckling.

The report will be viewed as a benchmark study in behavioural ecology and neuroscience. The authors combined expertise in investigation of the chemical senses, and in developmental psychobiology and analytical chemistry, and started by fractionating rabbit milk by gas chromatography. The behavioural potency of each of the resulting fractions was assessed by presenting them to pups through a port on the gas chromatograph or on a glass rod. The prelude to suckling involves a specific 'head-scanning' behaviour, which culminates in a pup's

identifying and locking on to the nipple. Schaal *et al.* found that pup responses to 2MB2 were indistinguishable from those induced by milk. Both elicited head scanning and, when 2MB2 was presented on a rod, seizing of the rod end; none of the 20 other fractions isolated from milk triggered the sequence.

The authors' further studies were characterized by their perspective and thoroughness. For example, in what will be seen as a tutorial on how to explore the intersection between analytical chemistry and behaviour, Schaal *et al.* showed an exactly parallel decline, due to compound volatility, in the effectiveness of milk and 2MB2 in eliciting responses: after 90 minutes of exposure to air, both had lost their behavioural potency. Applying the powerful logic of counter-experiment, favoured by the renowned nineteenth-century physiologist Claude Bernard, the authors found that when milk was deactivated it could no longer elicit the stereotyped responses, but that those responses were reinstated by adding 2MB2 at the concentration found in rabbit milk.

Other experiments showed that fresh milk or colostrum from several different species — including cattle, sheep, rodents, humans and horses — failed to elicit the stereotyped nipple-seeking behaviour in the breed of rabbit used. Conversely, 2MB2 elicited such patterns in six other breeds, but was not effective in a distant relative of rabbits, or in rodents or kittens. The authors

also provide strong evidence that 2MB2 is of mammary origin, in that it is present in milk but not blood plasma or amniotic fluid. The power of 2MB2 to elicit suckling did not depend on the mother's diet during or after pregnancy: pups of does that had been fed on totally different foods during pregnancy and after delivery reacted identically. Moreover, 2MB2 successfully elicited the initial suckling in rabbits that had been delivered by Caesarean section and those that had been delivered vaginally and captured before coming into contact with the mother's nipple region. This is further evidence that the effects of 2MB2 are not dependent on experience — that is, they are not learned.

Rabbits have an unusual mothering style, in that they visit and nurse their young for 5–7 minutes in a single daily visit⁴. This differs markedly from almost all other commonly studied mammals. Rat mothers, for example, are essentially in continuous contact with their young during the first week after delivery, leaving the nest only very briefly to eat and drink⁵. They also differ in being omnivores, not herbivores. The initial nipple attachment of rat young is controlled by their matching the odours of amniotic fluid detected prenatally with those detected postnatally⁶. The olfactory properties of the amniotic fluid reflect the mother's diet, entering the circulation and being filtered through the blood plasma during the fluid's formation. In rabbits, nipple attachment was not influenced by substances that had probably filtered into these fluids, even though pups can sense and respond to odours that are detected in amniotic fluid or milk.

A further difference between the species is that, whereas in rats the elicitors of suckling change, in rabbits 2MB2 continued to elicit nipple attachment at four days of age. In rats, by this stage, pup behaviour is probably based on the olfactory properties of the maternal diets, which the pups detect through the mother's milk (also a plasma filtrate)⁷. The same holds for humans — who, like rats, are omnivores, and also have protracted contact between the mother and infant⁸. For both species maternal diet, through amniotic fluid and milk filtrates, influences diet choice at the time of weaning and attraction to the mother⁹.

In their research with rabbits, Schaal *et al.*² have provided an anchor for future analyses of chemosensory influences on the development of animals with different feeding styles, reproductive strategies and 'nest history'. More immediate questions are not only how 2MB2 elicits the first attachment of pups, but how it signals the daily pattern of meals, given that milk instability and 2MB2 volatility render them ineffective within an hour, and that a doe visits her young only once daily. Perhaps a doe detects the activity of her pups as she approaches the subterranean nest through its connecting tunnel,

and perhaps — as occurs in human mothers in different circumstances — a modest leakage of milk is triggered that elicits the sequence reported by Schaal and co-workers.

Finally, here is a topic that invites the further involvement of neuroscientists. Given the possibility of identifying the configurations of protein receptors in the olfactory mucosa¹⁰, and of tracing the nerve inputs into the olfactory bulb and cortex in the brain, it should be possible to trace the neural pathway between the detection of 2MB2 and nipple attachment. And as the behavioural analyses progress, so too can the neurology. In particular, the linkage of food odours with 2MB2 might provide an understanding of post-weaning diet choice in rabbits, and might advance similar studies in animals with different life histories, including humans⁵. ■

Elliott M. Blass is in the Department of Pediatrics,

Boston Medical Center, Boston, Massachusetts, and the Department of Psychology and the Division of Neuroscience and Behavior, University of Massachusetts, Amherst, Massachusetts 01003, USA. e-mail: ebllass@medusa.sbs.umass.edu

1. Blass, E. M. & Teicher, M. H. *Science* **210**, 15–22 (1980).
2. Schaal, B. *et al.* *Nature* **424**, 68–72 (2003).
3. Beauchamp, G. K., Doty, R. L., Moulton, D. G. & Mugford, R. A. in *Olfaction, Behavior, and Mammalian Reproduction* (ed. Doty, R. L.) 143–160 (Academic, New York, 1976).
4. Coureaud, G., Schaal, B., Hudson, R. & Orgeur, P. *Dev. Psychobiol.* **40**, 372–390 (2002).
5. Johnson, B. A. & Leon, M. in *Handbook of Behavioral Neurobiology: Developmental Psychobiology* Vol. 13 (ed. Blass, E. M.) 53–80 (Kluwer/Plenum, New York, 2001).
6. Pedersen, P. E. & Blass, E. M. *Dev. Psychobiol.* **15**, 349–355 (1982).
7. Pedersen, P. E., Williams, C. L. & Blass, E. M. *J. Exp. Psychol.: Anim. Behav. Processes* **8**, 329–341 (1982).
8. Menella, J. A., Johnson, A. & Beauchamp, G. K. *Chem. Senses* **20**, 207–209 (1995).
9. Galef, B. G. Jr & Henderson, P. W. J. *Comp. Physiol. Psychol.* **78**, 213–219 (1972).
10. Buck, L. B. & Axel, R. *Cell* **65**, 175–187 (1991).

Ecology

Roots of diversity

Peter D. Moore

Competition between plants is in part responsible for the diversity of vegetation in different ecosystems. Diversity studies have to take into account what is happening beneath the soil as well as above it.

The subterranean life of plants is not easily documented. Take, for instance, the question of how root activities may determine the composition and diversity of plant communities. Above ground there is an observable struggle for light, but what goes on in the soil may be of greater consequence in determining diversity. Such is the conclusion reached by Tara Rajaniemi and colleagues from experiments that they describe in the *Journal of Ecology*¹.

Understanding the factors that control diversity is high on the agenda for ecologists and conservationists, and different factors evidently operate at different scales. For example, at a continental scale the areas that have the greatest diversity of species are often those with the highest levels of primary production. The productive forests of the equatorial regions support more species than do less productive systems in the cool temperate zone. In North America, the areas that are richest in tree species correspond to those regions (such as the south-east) where forest productivity is highest². But at the habitat scale this correlation does not apply. A reed bed may be highly productive, but it is relatively non-diverse, and productive grasslands are usually lower in plant diversity than less fertile ones³. Competition seems to lie at the heart of this problem, with robust and productive plants eliminating those that are less able to acquire the resources that they need. The most obvious of these are sunlight, nutrient elements and water, the

first being tapped by shoots, the latter two by roots.

Ecological research tends to concentrate on above-ground processes, because they are more straightforward to study. Rajaniemi and colleagues¹, however, have attempted to manipulate shoot and root competition separately so as to analyse the influences on ecosystem diversity. They conducted their experiments in an 'old field' habitat in Michigan, selecting seven of the ten most dominant plant species — including two grasses and five forbs (broad-leaved herbs). They set up plots in which each of these species was grown in monoculture and exposed to or protected from either the shoots or roots of other species. Protection was achieved by using nets to exclude shading from neighbouring shoots, and trenches to eliminate root competition from surrounding plants. Control plots were exposed to both shoot and root competition. Half of the plots were then fertilized to boost their productivity.

Thus, each of the seven species was subjected to either root or shoot competition from other species while in monoculture, or exposed to both of these in the control plots, and each treatment had a replicate that had been given fertilizer. Maintaining shoots, yet restraining them with nets, means that root systems below ground remain intact and active even though the shoots above ground are not responsible for any shading of the test plants.

The first conclusion from the intact plot experiments is that fertilization does indeed reduce plant diversity, as might be predicted for this scale of study. However, when Rajaniemi *et al.* examined the relative impacts of shoot competition and root competition on diversity, they found that virtually all the depression in diversity is a consequence of the effects of roots: the long-held assumption that shoot growth and competition for light lead to exclusion of species when productivity increases is not supported by this work. It could be that this result is related to the generally low stature of the vegetation (grassland) under examination here: perhaps competition for light is not critical at these biomass levels and in this relatively simply structured system. But the implication is that resources in the soil (in this case nutrient elements, because water is not limiting) are the subject of the most intense competition and that some species are more effective at tapping these resources than are others, thus achieving dominance.

Unfortunately, this study, like all others of its kind, fails to establish the extent and the allocation of root biomass between species, although this is not surprising given the difficulty in determining the precise identity of fine roots in the soil. One must assume that the most successful root competitor is the

plant with the most root tissue⁴ and the one that can expand its root biomass rapidly and effectively forage for the richer patches of nutrients within the soil.

In the real world of grazed pastures, the question of patchiness of resources (such as areas of dung deposition) and of disturbance (by intensive grazing or trampling) makes interpretations even more difficult⁵. If the strong competitor can rapidly extend its roots into the nutrient hotspots, then plants with smaller root systems or slower root growth can effectively be eliminated, thus reducing diversity. Rajaniemi and colleagues' study is a reminder that ecologists should pay attention to the world below the ground, where the further complications of assessing the effects of microbial activity and fungal symbionts also await them. ■

Peter D. Moore is in the Division of Life Sciences, King's College London, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NN, UK. e-mail: peter.moore@kcl.ac.uk

1. Rajaniemi, T. K., Allison, V. J. & Goldberg, D. E. *J. Ecol.* **91**, 407–416 (2003).
2. Currie, D. J. & Paquin, V. *Nature* **329**, 326–327 (1987).
3. Grime, J. P. *Plant Strategies, Vegetation Processes, and Ecosystem Properties* (Wiley, Chichester, 2001).
4. Newman, E. I. *Nature* **244**, 310–311 (1973).
5. Kemp, D. R. & King, W. McG. in *Competition and Succession in Pastures* (eds Tow, P. G. & Lazenby, A.) 85–102 (CABI, Wallingford, 2001).

Astronomy

Heartbeats of a neutron star

Robert V. Wagoner

Different oscillations in the flux of X-rays emitted from a neutron star now seem to be linked to the frequency at which the star is spinning. That spin rate in turn hints at what the interior of such a star is made of.

Neutron stars are remnants of the supernova explosions of stars more massive than our Sun. Their dense cores are thought to consist mostly of neutrons, but may contain exotic states of matter that are unknown elsewhere in the Universe. Neutron stars are often found in binary systems, partnered by a surviving star. Hot gas spirals onto the neutron star from its stellar companion, dragged in by the gravity of the denser object, and X-rays are emitted (Fig. 1, overleaf). Observations by the Rossi X-ray Timing Explorer (RXTE) satellite and other spacecraft during the past decade have shown that these X-rays carry three types of nearly periodic, high-frequency signals¹. Now, as reported on pages 42 and 44 of this issue^{2,3}, all three types of signal have been detected from a single neutron-star binary, known as SAX J1808.4–3658. The observations point to the rate of rotation of the neutron star as a critical factor in the generation of these signals.

Chakrabarty *et al.*² provide strong evidence that the final frequency of brightness

oscillations during bursts of X-rays (thought to be produced by thermonuclear explosions on the surface of the neutron star) and the frequency of another persistent, coherent signal can both be identified with the rate of rotation, or spin frequency, of the neutron star. Wijnands *et al.*³ report the discovery of twin quasi-periodic oscillations (QPOs) in the kilohertz range, and the difference between their frequencies is very close to half the spin frequency. Such QPOs are believed to be produced within the gas orbiting above the neutron star's surface, but no models have predicted this relationship to its spin.

These observations provide important probes of the extreme conditions inside neutron stars and of the strong gravitational fields that they generate. The tidal force exerted by a neutron star on its companion star induces a flow of matter from the star through a rotating accretion disk onto the neutron star (Fig. 1). As a result, the spin frequency of the neutron star increases, an effect known as 'spin-up'. If, eventually, the



100 YEARS AGO

Recently during the course of an experiment I had occasion to boil water in presence of mercury. After ebullition had been going on for some time, I noticed that occasionally steam forming below the surface of the mercury carried with it a pellicle of mercury as it rose through the water in the form of a bubble. When it reached the surface of the water the pellicle usually burst, and the mercury fell back as a drop. By adjusting the intensity of ebullition, it was possible to bring the two liquids into such a state that, comparatively frequently — say ten times per minute — steam bubbles covered with mercury rose through the water and floated on its surface, and, hovering there for an instant, they cooled and contracted, and sank slowly down through the water. When the bubbles are formed in rapid succession, the phenomenon is one of great beauty. From *Nature* 2 July 1903.

50 YEARS AGO

A number of clinical observations have established that recent immunization with diphtheria or whooping cough vaccines renders children more susceptible to poliomyelitis infection... A similar effect has been demonstrated in mice infected with the virus of mouse encephalomyelitis (GDVII strain) in experiments which are being reported fully elsewhere. During the course of the work, however, it was observed that the injection of vaccines which had been stored for several months no longer made the mice more susceptible to virus infection, but instead exerted a protective effect. It is felt that these findings, if confirmed, might have an important bearing on both clinical and biological problems.

ALSO...

To mark the centenary of the birth of Sir Henry Wellcome on August 21, 1853, a commemorative exhibition will be opened at the Wellcome Foundation, Ltd., on Wednesday, July 8, by Mr. Winthrop W. Aldrich, American Ambassador in London. This exhibition... will cover the whole period of Sir Henry's long life (1853–1936). The exhibits will range from interesting relics of his early days in the United States to sections on some of the world-famous research institutes which he established after he settled in Great Britain: the Wellcome Physiological Research Laboratories, Wellcome Research Institution, Wellcome Museum of Medical Science, and Wellcome Historical Medical Museum and Library. From *Nature* 4 July 1953.

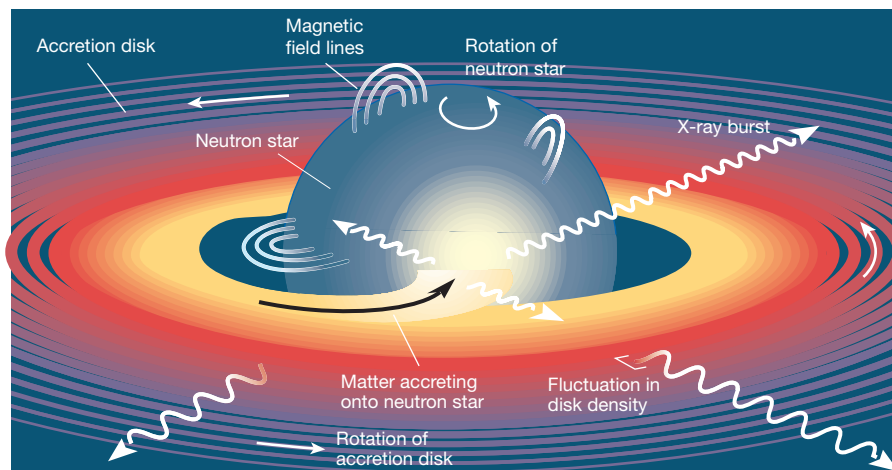


Figure 1 A neutron star and its inner accretion disk. The gas comprising the disk orbits around and gradually spirals onto the neutron star. The star's magnetic field can modify the flow of gas onto its surface. Two sites of X-ray modulation are shown: a thermonuclear X-ray burst spreading on the neutron star, and a fluctuation near the inner edge of the accretion disk. Chakrabarty *et al.*² and Wijnands *et al.*³ have detected three different X-ray signals from a single neutron star, and all three seem to be linked to the star's spin frequency.

companion is fully consumed, the neutron star becomes an isolated 'pulsar', emitting at radio (and shorter) wavelengths.

The maximum rate at which a neutron star can spin (without shedding mass from its equator) depends on its composition and physical state. General relativistic calculations⁴ show that spin-up of slowly rotating neutron stars stops when their spin frequency reaches at least 1,180 Hz. As Chakrabarty *et al.*² point out, the distribution of spin frequencies of accreting neutron stars is known to be concentrated near a maximum of 619 Hz. The fact that these values are different suggests that some mechanism prevents further spin-up.

There are two plausible explanations. The first is that the neutron star's magnetic field is strong enough to push out the inner radius of the accretion disk and eliminate the gas of highest orbital frequency. However, this magnetic field should also channel the X-ray-emitting gas to preferred locations on the surface, whose rotation should produce a coherent oscillating X-ray component. Although Chakrabarty *et al.* did observe such a coherent component, it has not been seen in the ten other neutron star X-ray sources that also exhibit burst oscillations. Alternatively, density or velocity fluctuations within the neutron star might produce gravitational radiation, reducing its angular momentum and hence limiting the spin frequency. This gravitational radiation might be detected by the (Earth-based) kilometre-scale laser interferometers currently being developed in the United States, Europe and Japan⁵.

Of the many X-ray sources seen to contain twin high-frequency QPOs, seven were also seen to produce X-ray burst oscillations (but no coherent component). Each QPO spans a narrow range of frequencies, and the average frequency usually varies significantly

with the rate at which the neutron star is accreting mass. As also reported by Wijnands *et al.*³, often one (or both) of the QPOs may disappear for some time. Although the frequency difference of the pair remains fairly constant, in some of these sources it is close to the X-ray burst oscillation frequency (now identified with the spin frequency), but in the others it is close to half that (as in SAX J1808.4–3658). This seems to require that the spin of the neutron star be coupled (via magnetic fields or radiation from the surface) to the gas in the accretion disk, where the QPOs are thought to originate.

Density and temperature inhomogeneities in the accretion disk presumably generate the QPOs. If there are concentrations, or blobs, of gas within the disk, they could be eclipsed by the neutron star, causing a variation of the X-ray flux; and as the blob velocity along the line of sight varies, this also causes a variation of the radiation intensity through the Doppler effect. Alternatively, the

inhomogeneities can be produced by coherent waves (normal modes) trapped within certain regions of the disk. The interaction of magnetic fields and/or radiation from the neutron star's surface with orbiting blobs could couple the star's spin to the orbital motion of the disk material⁶. However, as Wijnands *et al.*³ point out, the most general such model, involving both surface and accretion-disk inhomogeneities, cannot produce a QPO frequency difference that is half the spin frequency (although the simplest model can explain those cases in which the QPO frequency difference and spin frequency are similar in value).

Wijnands *et al.*³ propose that the problem might be resolved by combining elements from other models — in particular, resonances that arise between the star's spin frequency and orbital precession frequencies (predicted by general relativity), which occur at specific radial distances in the accretion disk. But the implementation of this proposal remains a challenge, especially as our knowledge of the actual physical conditions in the environment of these neutron stars is so uncertain.

The results reported by Chakrabarty *et al.*² and Wijnands *et al.*³ have sharpened our perception of two fundamental mysteries: what stops the spin-up of neutron stars, and how their spin rate is communicated to their accretion disk. The answers to these questions might involve exotic physics, but, with more data from X-ray missions such as RXTE and perhaps from gravitational-wave detectors, they are hopefully not far away. ■

Robert V. Wagoner is in the Department of Physics, Stanford University, Stanford, California 94305-4060, USA.

e-mail: wagoner@stanford.edu

1. van der Klis, M. *Annu. Rev. Astron. Astrophys.* **38**, 717–760 (2000).
2. Chakrabarty, D. *et al. Nature* **424**, 42–44 (2003).
3. Wijnands, R. *et al. Nature* **424**, 44–47 (2003).
4. Cook, G. B. *et al. Astrophys. J.* **423**, L117–L120 (1994).
5. Cutler, C. & Thorne, K. S. in *Proc. 16th Int. Conf. General Relativity and Gravitation* (eds Bishop, N. T. & Maharaj, S. D.) (World Scientific, Singapore, 2002).
6. Miller, C. M. *et al. Astrophys. J.* **508**, 791–830 (1998).

Ion channels

Hearing aid

Rachel A. Dumont and Peter G. Gillespie

Mechanically controlled ion channels — transduction channels — are a key feature of the cells that detect sound, touch and movement. In fruitfly ears, the channels belong to a very familiar group of proteins.

Sensory information, such as that carried by light, odours and sound, continuously washes over us, revealing the nature of the outside world. Because organisms place a premium on detecting such signals with high specificity and sensitivity, specialized cells have evolved to capture each type of stimulus. Mechanical

stimuli are no exception, and most organisms use mechanoreceptor cells to detect sound, touch and movement. Central to mechanoreception are ion channels located in the outer membranes of mechanoreceptor cells; these 'transduction' channels detect mechanical stimuli and, by modulating ion entry, control the cells' electrical

excitability in proportion to the size of the stimulus. That excitation in turn enables appropriate electrical signals to be sent to the brain. The precise channels involved in mechanoreception have, for the most part, evaded identification, because the cells are exceedingly scarce. However, in an exciting report on page 81 of this issue, Kim and colleagues¹ show that Nanchung — a newly discovered member of the TRP protein family — is the transduction channel needed for fruitflies to hear.

A fly detects courtship songs with its ear, Johnston's organ². The molecular details of this feat are largely unknown, however. One reason for wanting to find out is that many molecular systems in *Drosophila* are closely related evolutionarily to those in higher organisms — and so identifying the molecules that transduce mechanical stimuli in flies should have broader significance. Advantages to studying mechanotransduction in flies rather than more complex organisms include the ability to carry out genetic screens.

Although several such screens have been conducted, progress in finding the molecules responsible for touch perception and hearing has been slow³. A notable success was the discovery that flies rely on NompC, another TRP channel, to detect movement of their sensory bristles — small hairs that decorate their bodies⁴. NompC could not, however, be the only fly transduction channel; fruitflies that lack this protein still retain a small residual electrical response when their bristles are deflected⁴. Moreover, these flies enjoy nearly normal hearing². Additional transduction channels clearly awaited discovery.

Nanchung ('Nan' for short) is one such channel: Kim *et al.*¹ present compelling evidence that this TRP-family member is essential for *Drosophila* hearing. In the absence of the *nan* gene, flies do not hear. Moreover, *nan* is expressed by mechanoreceptor cells in Johnston's organ, particularly in the outer dendritic segment (Fig. 1), where transduction is thought to take place. Finally, expression of *nan* in tissue-culture cells leads to the appearance of a channel that is activated by one type of mechanical stimulus, osmotic challenge. Although Nan might be working by activating another channel in these cells, it is more likely that Kim *et al.* have described some of the conductance properties of the transduction channel that is directly responsible for sound detection. TRP channels — named after *transient receptor potential*, a fly gene essential for phototransduction — have been found to mediate sensory transduction in a wide variety of cell types⁵. Nan belongs to the TRPV subfamily, which has been implicated in sensing pain, temperature, odours, osmotic pressure and touch⁶. With Kim and colleagues' findings, we can now add sound to this list.

Fly hearing may be a valuable model system for understanding other mechanoreceptors, not least the vertebrate inner ear's

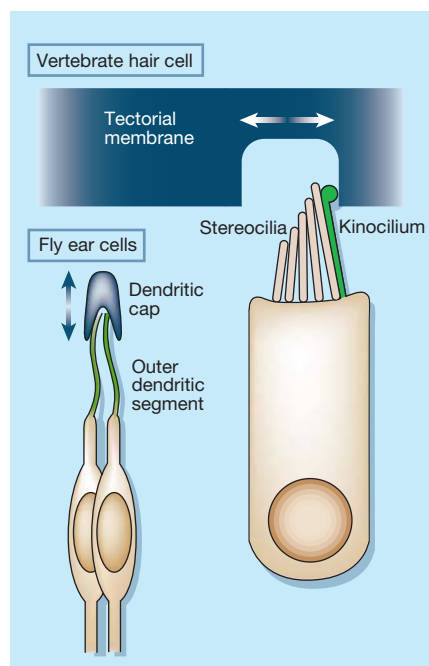


Figure 1 Fly and vertebrate auditory mechanoreceptors. In the fly ear, pairs of mechanoreceptor cells are stimulated when the overlying dendritic cap, in response to sound waves, moves relative to the cells (arrow). Kim *et al.*¹ have found that the Nan ion channel is located in the outer dendritic segment, where transduction — the detection of sound stimuli and their conversion into the electrical response of the cell — is thought to take place. In a vertebrate hair cell, the movement of the overlying tectorial membrane relative to the cell deflects stereocilia, triggering transduction; the kinocilium helps to couple these processes. The kinocilium is present in hair cells in all vertebrates, at least during development. Structures that are anatomically and molecularly similar between fly ear cells and vertebrate hair cells include the dendritic cap and tectorial membrane (blue), and the outer dendritic segment and kinocilium (green).

hair cell. This hypothesis has not been accepted by all, perhaps in part because of differences in appearance between the two cell types (Fig. 1). For instance, fly mechanoreceptors such as those in Johnston's organ rely on a single long, thin protrusion, or 'sensory process', to perceive mechanical stimuli⁷. By contrast, a vertebrate hair cell carries out transduction with its hair bundle, which is composed of some 100 clustered processes called stereocilia⁸. Beyond sheer numbers, these processes differ fundamentally in their structural proteins: hair-cell stereocilia are filled with actin, whereas the fly sensory process derives from a cilium constructed of tubulin.

Despite these anatomical differences, there are some remarkable similarities. Both types of sensory process are bathed in an unusual fluid, endolymph, which is distinguished by a high K^+ concentration and a substantial extracellular potential^{7,8}. Moreover,

the responses of *Drosophila* and vertebrate mechanoreceptors to mechanical stimuli show striking resemblances in their speed, sensitivity and adaptation⁴. Finally, vertebrate and invertebrate ears have conserved the molecules that control development of mechanoreceptor cells⁸. These cells may well be derived from a common, ancient precursor — supporting that idea, although vertebrate hair cells rely on stereocilia for transduction, they too have a tubulin-based cilium (the kinocilium; Fig. 1).

If these cells are evolutionarily linked, additional molecular similarities should emerge — and indeed they have. In one example, NompA, a component of the structure that caps the fly's sensory process⁹, resembles the tectorins — proteins that contribute to a similar structure in vertebrate ears (Fig. 1). In another, NompC (found in *Drosophila* bristles) underlies transduction in zebrafish hair cells¹⁰. Although mammalian genomes conspicuously lack a close relative of the *nompC* gene, its presence in zebrafish nevertheless dramatically reinforces the hypothesis that the molecules responsible for *Drosophila* mechanotransduction have relatives in vertebrate systems.

Could the transduction channel in mammalian hair cells — like that in fruitfly ears¹ — belong to the TRPV subfamily? Of the six known vertebrate TRPV proteins, a tempting candidate is TRPV4, which is found in one class of hair cells (auditory hair cells) and can be activated osmotically when expressed in tissue-culture cells¹¹. To complicate the story, however, the hair-cell transduction channel may not be made up of a single type of protein; for example, multiple TRP-channel subunits interact to achieve correct targeting and function of sensory channels in the worm *Caenorhabditis elegans*¹². Although the demonstration of the role of Nan in fly hearing does not pinpoint the identity of the hair-cell transduction channel, it does direct our gaze firmly at the TRPV subfamily.

Rachel A. Dumont and Peter G. Gillespie are at the Oregon Hearing Research Center and the Vollum Institute, Oregon Health & Science University, Portland, Oregon 97239, USA.

e-mail: gillespp@ohsu.edu

- Kim, J. *et al.* *Nature* **424**, 81–84 (2003).
- Eberl, D. F., Hardy, R. W. & Kernan, M. J. *J. Neurosci.* **20**, 5981–5988 (2000).
- Ernst, G. G. & Chalfie, M. *Annu. Rev. Genet.* **36**, 411–453 (2002).
- Walker, R. G., Willingham, A. T. & Zuker, C. S. *Science* **287**, 2229–2234 (2000).
- Montell, C., Birnbaumer, L. & Flockerzi, V. *Cell* **108**, 595–598 (2002).
- Gunthorpe, M. J., Benham, C. D., Randall, A. & Davis, J. B. *Trends Pharmacol. Sci.* **23**, 183–191 (2002).
- Keil, T. A. *Microsc. Res. Tech.* **39**, 506–531 (1997).
- Gillespie, P. G. & Walker, R. G. *Nature* **413**, 194–202 (2001).
- Chung, Y. D., Zhu, J., Han, Y. & Kernan, M. J. *Neuron* **29**, 415–428 (2001).
- Sidi, S., Friedrich, R. W. & Nicolson, T. *Science* published online 12 June, doi:10.1126/science.1084370 (2003).
- Liedtke, W. *et al.* *Cell* **103**, 525–535 (2000).
- Tobin, D. *et al.* *Neuron* **35**, 307–318 (2002).

Obituary

Ilya Prigogine (1917–2003)

Ilya Prigogine died on 28 May in Brussels, after a long illness. Born in Moscow, he emigrated at an early age, as his family sought to escape the aftermath of the Bolshevik revolution. The family arrived first in Germany, then moved to Belgium — the country that became Prigogine's true homeland. He studied chemistry and physics at the Université Libre de Bruxelles, and obtained his PhD in 1939 under the tutelage of Théophile De Donder, a remarkable professor, who pioneered the development of modern physics in Belgium. Extending and deepening De Donder's work, Prigogine himself became a pioneer of the thermodynamics of irreversible processes.

Classical thermodynamics is the science of equilibrium: the concept of variation with time was viewed with suspicion; dissipative phenomena, such as friction, were considered a nuisance. Prigogine attacked the problem head-on, by introducing a quantitative concept of irreversibility. He produced a sound derivation of transport processes, relating the fluxes of energy and matter to the thermodynamic forces (such as gradients of temperature or density, or electric fields) that cause them. Irreversibility and entropy became the main themes of his subsequent work.

The next step was to consolidate these macroscopic notions by addressing their molecular basis. This subject had remained in the shadows since the last years of the nineteenth century with Ludwig Boltzmann's work on the kinetic theory of gases and a definition of entropy at the microscopic level. (It was even considered to be 'cursed' — Boltzmann committed suicide in 1906.) How could the irreversibility of real processes at the macroscopic scale be reconciled with the perfect time-reversibility of the (classical or quantum) law of motion for molecules? In the 1960s, surrounded by a small group of enthusiastic co-workers, Prigogine began to make crucial progress on this point, developing the first form of non-equilibrium statistical mechanics. (Another approach, developed independently by Nikolai Bogolyubov in the Soviet Union, turned out to be equivalent to Prigogine's.) As well as being quite a general formulation of the theory, there were other fruits of this endeavour, for it could be applied to a variety of systems, from gases and solids to plasmas.

Pushing the study further, to systems



Pioneer of the thermodynamics of irreversible processes

that are far from equilibrium, Prigogine and his increasingly numerous co-workers uncovered an extraordinary phenomenon. When the distance from equilibrium (measured by some appropriate parameter) reaches a certain threshold, the trajectory of a system meets a fork, or bifurcation. The system may then leave the trajectory along which it has evolved from equilibrium, and jump to a totally different one. If the system is pushed even further, more bifurcations may appear, and when the system is very far from equilibrium, it may behave completely chaotically (as, for example, the regular flow of a liquid may become turbulent). But, alternatively and unexpectedly, the system might reach a new, ordered state — what Prigogine called a 'dissipative structure'.

Such states are particularly striking in systems in which chemical reactions proceed alongside the effects of diffusion or external forces: ordered structures of different chemical composition appear, which may even propagate as 'chemical waves'. Two ingredients are vital if this is to occur: the 'open' character of the system (meaning that it can exchange matter and energy with the external world), and the nonlinear character of the equations

governing the evolution of the system. These conditions arise, in particular, in living systems. Prigogine had thus created an important link between physics, chemistry and biology (even extending it to sociology and economics). His achievement was crowned with the award of the 1977 Nobel prize for chemistry — the pinnacle of a long list of awards, prizes and honorary degrees.

By the 1990s, Prigogine was again thinking about physics at the microscopic level. Substantial progress had been made by mathematicians and physicists in the study of nonlinear dynamical systems, and one of the most important features to be uncovered was the intrinsic dynamical instability of most 'non-integrable' systems (even very simple ones). As a result, Prigogine introduced the crucial idea that the usual method of defining the state of a system by specifying exactly the positions and momenta of all of its components (that is, a point in 'phase space') is not realistic, because a close, neighbouring state may evolve in a completely different way. Rather, the state should be described by an ensemble — a cluster of identical systems differing in their initial conditions. The latter may (but does not have to) be concentrated around a single point in phase space. The description of the system's evolution thus becomes statistical. In this way, the solutions of the equations are regularized (divergences are suppressed), and irreversibility appears as a welcome bonus in the theory.

Like his mentor De Donder before him, Prigogine was also a remarkable professor. In the United States, he was the founder and director of the Center for Statistical Mechanics at the University of Texas at Austin (later renamed the Ilya Prigogine Center for Studies in Statistical Mechanics and Complex Systems); in 1959 he was made director of the International Solvay Institutes for Physics and Chemistry in Brussels. His lectures were fascinating for students, as he preferred to leave out tedious details and instead include parenthetical perspectives on art, music and philosophy. His books for a general audience, such as *La Nouvelle Alliance* (with I. Stengers), *From Being to Becoming* and his last work, *La Fin des Certitudes*, were bestsellers around the world. He was a true humanist, in the widest meaning of the word, and attracted numerous disciples. His death closes an important chapter in the history of science.

Radu Balescu

Radu Balescu is at 86 Avenue A. Huysmans, B-1050 Bruxelles, Belgium.
e-mail: rbalescu@tiscalinet.be

Plant biology

Blooming extension at the roadside

J. Evol. Biol. **16**, 551–557 (2003)

Two plant species that do not normally interbreed do so in habitats that are subject to human disturbance. So conclude B. B. Lamont and colleagues from their study of the shrub *Banksia hookeriana* and the closely related tree *B. prionotes*, two sand-dune-hugging members of the scrub-heath flora of Western Australia.

These species share a common pollinator and hybridize readily when artificially pollinated. But natural hybrids are rare because their flowering seasons do not overlap. Lamont *et al.* found, however, that the plants' biology in disturbed environments beside roads and railway lines differs from that of plants in undisturbed areas, and that hybrids do arise there. In such environments, individuals of both species grow taller and produce more flowers, apparently because of more water and soil nutrients. Crucially, the flowering seasons extend at both ends so that they overlap, breaking the barrier to interbreeding.

The genetic integrity of parent plant species may be compromised if the hybrid offspring have a competitive advantage over them. So the route to hybridization described by Lamont and colleagues is one for conservation biologists to keep an eye on.

Philip Fletcher

Atmospheric chemistry

Water molecules pair in air

Science **300**, 2078–2080 (2003)

Water molecules may pair up in the atmosphere, say K. Pfeilsticker and colleagues. They have detected hitherto elusive water dimers in ambient air.

In liquid water, the molecules stick to one another via hydrogen bonds. They may be bound into small clusters in the same way in the gas phase, but the concentration of such clusters was thought to be very small in the atmosphere. Previous attempts to find water dimers in air have failed.

The (near-infrared) spectral signature of water dimers is well known from laboratory studies, and Pfeilsticker and colleagues were able to detect it by measuring absorption by the dimers along an 18.34-km path through the lowest region of the atmosphere (troposphere). They estimate that one in every 1,000 water molecules in vapour-saturated air at 20 °C is bound in a dimer.

Larger clusters are also predicted to be stable in the atmosphere, but the researchers conclude that the concentration of such clusters decreases by a factor of around 1,000 for every extra molecule added to a cluster. All the same, this implies that clusters with

up to six molecules exist in the troposphere in numbers comparable to the ambient concentration of aerosol particles. Here they can facilitate photochemical reactions such as the formation of sulphuric acid, 'seed' the formation of water droplets, hydrate other atmospheric gases, and alter the atmosphere's radiative balance.

Philip Ball

Cancer

The alcohol factor

Cancer Res. **63**, 3092–3100 (2003)

As cells break down the ethanol in alcoholic drinks, the by-products can contribute to DNA damage and so to cells becoming cancerous. Elise A. Triano *et al.* have now found that normal human breast epithelial cells contain the enzyme class I alcohol dehydrogenase (ADH), which can metabolize alcohol. They propose that this might help to protect breast-feeding infants from alcohol's toxic effects — but it might also be responsible, in part, for the observation that drinking alcohol increases the risk of breast cancer.

The authors also found, however, that levels of ADH were greatly decreased or absent in cells taken from invasive breast cancers, leading them to suggest that the protein might also have some tumour-suppressor function. Class I ADH can affect a wide range of molecules besides ethanol. Most importantly, it helps to convert vitamin A to retinal — the first step in the production of retinoic acid, which has a restraining influence on cell division. So, in invasive cancer cells, this effect of retinoic acid might be compromised because of decreased expression of the enzymes needed to generate it from vitamin A.

Helen R. Pilcher

Biophysics

Symmetry by design

Phys. Rev. Lett. **90**, 248101 (2003)

Sphere-like viruses have icosahedral symmetry not by default but by design, say Robijn F. Bruinsma and colleagues. This undermines the common assumption that the shapes of viruses are dictated by simple geometric packing constraints on their protein building-blocks.

The protein coats (capsids) of viruses are made from individual, compact proteins grouped into geometric units such as hexagons, and known as capsomers. In 1956, Crick and Watson proposed that spherical viruses should have the symmetry of regular polyhedra because of the way small numbers of identical protein units pack together. Most such viruses are indeed icosahedral, suggesting that this shape is simply the one that minimizes the packing energy.

Not so, say Bruinsma and colleagues. Their model, which considers the adhesion and bending energies of closed shells of

disk-like capsomers, reveals that the most stable structures need not be icosahedral.

Instead, icosahedral shells can be explained by assuming that (as is often found in reality) 12 of the capsomers are smaller than the others. This can be engineered by a 'conformational switch' of the constituent proteins, which has an energy cost. In real viruses, this modification to the building-blocks seems to be an adaptation — although its function isn't clear.

Philip Ball

Immunology

Hide and seek

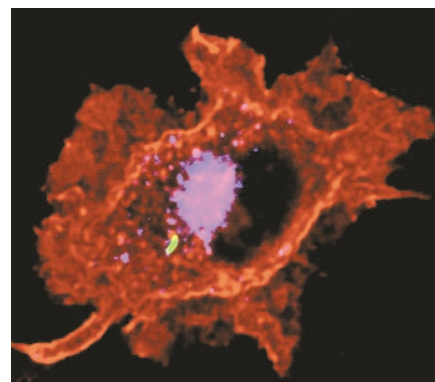
Immunity **18**, 813–823 (2003)

Legionella pneumophila, the bacterium that causes a severe type of pneumonia known as Legionnaire's disease, hides in a clandestine compartment within our cells. Annie L. Neild and Craig R. Roy have found that dendritic cells — experts in turning on the immune response — can nevertheless seek out the intruder.

Legionella multiplies vigorously within macrophage cells in the lungs. Normally, macrophages can present bits of bacterial proteins to cells of the immune system to elicit an immune attack. But *Legionella* lives in an atypical intracellular compartment, a vacuole, which never seems to mix with the compartment in which proteins are processed for presentation. Moreover, rapid replication of *Legionella* can kill a macrophage within 24 hours. These considerations make macrophages unlikely candidates for setting off an immune response to the bacterium — in theory.

Neild and Roy now show that *Legionella* resides in a vacuole in dendritic cells as well. And, surprisingly, *Legionella*-derived protein products are somehow presented to the immune system by both dendritic cells and macrophages *in vitro*. Unlike macrophages, however, dendritic cells manage to firmly restrict the growth of their unwelcome guest. How they do that is not known, but it may allow dendritic cells *in vivo* sufficient time to live and tell the tale — to the immune system at least.

Marie-Thérèse Heemels



Unwelcome guest: *Legionella* (green) inside a macrophage (red). Lysosomes are blue.

ELSEVIER

Insect orientation to polarized moonlight

An African dung beetle uses the moonlit sky to make a swift exit after finding food.

Moonlight, like sunlight¹, scatters when it strikes tiny particles in the atmosphere, giving rise to celestial polarization patterns². Here we show that an African dung beetle, *Scarabaeus zambesianus*, uses the polarization of a moonlit sky to orientate itself so that it can move along a straight line. Many creatures use the Sun's light-polarization pattern to orientate themselves^{3,4}, but *S. zambesianus* is the first animal known to use the million-times dimmer polarization of moonlight for this purpose.

S. zambesianus starts to forage on the wing for fresh dung at around sunset. Once a source has been located, the beetle quickly

forms a ball of dung with its front legs and head (Fig. 1a), and rolls it away in a straight line — the most efficient path for escaping aggressive competition for food in the dung pile. To orientate itself to follow a straight line at dusk, *S. zambesianus* relies on the polarization pattern formed around the setting Sun to maintain its departure bearing⁵.

But after astronomical twilight, when the Sun is more than 18° below the horizon, this cue is no longer available. To find out whether the beetles are able to use the polarization of moonlight instead, we monitored their movements under the night-time sky. We found that on moonlit nights, the beetles rolled radially in straight lines away from the dung (Fig. 1b), but that on nights without a moon, they no longer followed a straight-line path (Fig. 1c).

To test whether the beetles' orientation depends on the polarization of the moonlit sky, rather than on the moon itself, we hid the moon from view by shading an arena that contained the beetle and its ball from the rising moon, and placed a polarizing filter over a ball-rolling beetle. The filter had its electric vector (*e*-vector) transmission axis orientated perpendicularly to the dominant orientation of the *e*-vector present in the zenith and along the lunar vertical, so the moon's polarized-light pattern appeared to turn through 90° as the beetles continued to roll beneath the filter. In response to this light rotation, the beetles turned close to the expected 90°, either left or right (Fig. 1d, e). The symmetrical pattern of polarized light from the sky does not allow the beetles to discriminate between left and right. As a control, the experiment was repeated with

the filter's *e*-vector orientation parallel to that of the sky, and this resulted in the beetle maintaining its rolling direction under the filter (Fig. 1e). We conclude that it was not the fivefold drop in light intensity experienced by beetles covered by a filter that caused them to change direction.

Our results indicate that these dung beetles orientate by using the polarization pattern of moonlight. Receptors for polarized-light analysis have recently been found in the dorsal-most part of the eye of *S. zambesianus*⁵. By using the polarization of moonlight for orientation, *S. zambesianus* is able to extend its foraging time. Although, to our knowledge, this is the first description of an animal using the moon's polarization pattern as a nocturnal compass, this ability may turn out to be widespread in the animal kingdom.

Marie Dacke*, Dan-Eric Nilsson*,
Clarke H. Scholtz†, Marcus Byrne‡,
Eric J. Warrant*

*Department of Cell and Organism Biology,
University of Lund, 223 62 Lund, Sweden
e-mail: marie.dacke@cob.lu.se

†Ecophysiological Studies Research Group,
Department of Animal, Plant and Environmental
Sciences, University of the Witwatersrand,
Wits 2050, South Africa

‡University of Pretoria, Department of Zoology and
Entomology, University of Pretoria, Pretoria 0001,
South Africa

1. Strutt, J. (Lord Rayleigh) *Phil. Mag.* **41**, 107–120 (1871).
2. Gál, J., Horváth, G., Barta, A. & Wehner, R. *J. Geophys. Res.* **106**, 22647–22653 (2001).
3. Waterman, T. H. *Handbook of Sensory Physiology* Vol. VII/6B (ed. Autrum, H.) 281–469 (Springer, Berlin, 1981).
4. Wehner, R. *J. Exp. Biol.* **204**, 2589–2596 (2001).
5. Dacke, M., Nordström, P. & Scholtz, C. H. *J. Exp. Biol.* **106**, 1535–1543 (2003).

Competing financial interests: declared none.

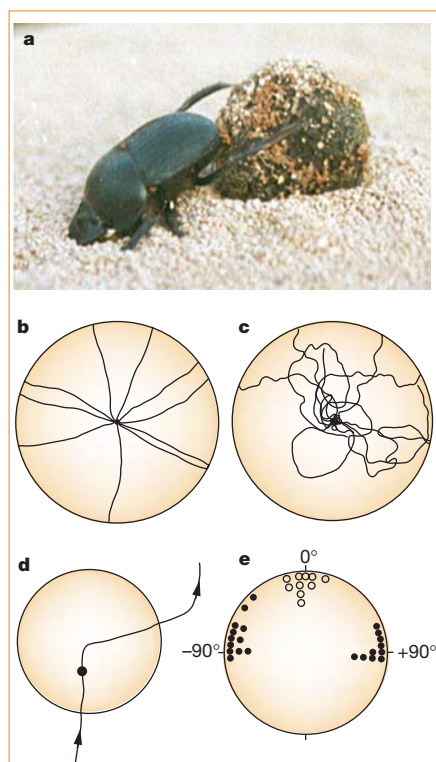


Figure 1 Polarization of moonlight and orientation in African dung beetles. **a**, The beetle rolling a ball of dung. **b**, **c**, Paths taken by beetles ($n = 10$) moving dung balls outwards from the centre of an arena (diameter, 3 m). On a moonlit night (**b**), beetles orientate along straight paths; on moonless nights (**c**), their direction is random. Maximum light intensities were $3.53 \times 10^{-2} \text{ cd m}^{-2}$ and $2.80 \times 10^{-4} \text{ cd m}^{-2}$ in **b** and **c**, respectively. **d**, Change in direction (turn to the right by $+70^\circ$) taken by a beetle when a perpendicularly polarizing filter (Polaroid HN22; circle represents the extent of the 42-cm-diameter filter) is placed over the beetle at the point indicated by the dot; the beetle resumes its direction of travel on exposure to the open sky. **e**, Average angles of turn made by 22 beetles when covered by the same filter as in **d** (filled circles; binned in 5° intervals): left turns, $n = 12$, $-77.0^\circ \pm 14.7^\circ$ (mean $\pm$ s.d.); right turns, $n = 10$, $+87.9^\circ \pm 9.3^\circ$. Under a parallel-polarizing filter ($n = 10$), beetles deviated by an average absolute angle of $6.7^\circ \pm 5.5^\circ$ (open circles; binned in 5° intervals) from the path they were following before filter placement.

COMMUNICATIONS ARISING

Photosynthesis

A new function for an old cytochrome?

In many cyanobacteria and algae, cytochrome c_6 transports electrons between the cytochrome *bf* complex and photosystem I, replacing plastocyanin when copper is deficient. Higher plants, however, were thought to lack cytochrome c_6 (refs 1,2) until the existence of a modified form in several species was inferred from genomic evidence³. By measuring oxygen evolution with inside-out thylakoids, Gupta *et al.* inferred that heterologously expressed *Arabidopsis* cytochrome c_6 can replace plastocyanin from *Synechocystis* or *Arabidopsis* in reconstitution experiments

*in vitro*⁴. From structural and kinetic evidence, however, we find that *Arabidopsis* cytochrome c_6 cannot carry out the same function as *Arabidopsis* plastocyanin or as cytochrome c_6 from the alga *Monoraphidium braunii*. This suggests that cytochrome c_6 in higher plants may have lost its original primitive function in photosynthesis.

The ultraviolet/visible-light absorption spectrum of reduced *Arabidopsis* cytochrome c_6 expressed in *Escherichia coli* is similar to that of *Monoraphidium* cytochrome c_6 , except that the absorption bands of the plant cytochrome are shifted towards the red (α -peaks are at 554.5 and 552.5 nm for plant and algal cytochromes, respectively; results not shown).

As expected, the midpoint redox potential of *Arabidopsis* plastocyanin (365 mV) is

similar to that of *Monoraphidium* cytochrome c_6 (358 mV), but surprisingly, that of *Arabidopsis* cytochrome c_6 (140 mV) is lower. *Arabidopsis* cytochrome c_6 would therefore be thermodynamically unsuitable for oxidizing cytochrome f (redox potential, 320 mV; ref. 5), but could, in principle, donate electrons to photosystem I. However, kinetic analysis of laser-flash-induced reduction of *Arabidopsis* photosystem I reveals that *Arabidopsis* cytochrome c_6 is 100 times less effective than plastocyanin as an electron donor (Fig. 1a), whereas algal cytochrome c_6 and *Arabidopsis* plastocyanin react at similar rates. The traces for *Arabidopsis* plastocyanin and *Monoraphidium* cytochrome c_6 are both biphasic, as is typical in eukaryotic systems⁶.

Figure 1b shows that the observed rate constant (k_{obs}) for the overall reaction with *Arabidopsis* cytochrome c_6 does not significantly increase even at high concentrations of electron-donor protein, whereas k_{obs} increases to a plateau for the slow phase with *Arabidopsis* plastocyanin and *Monoraphidium* cytochrome c_6 .

The dependence on ionic strength of k_{obs} clarifies the lack of reactivity of *Arabidopsis* cytochrome c_6 towards photosystem I (Fig. 1c). If this were due to charge repulsion between the donor and acceptor sites on the two proteins, k_{obs} should increase with ionic strength as a result of charge screening. This was not the case. By contrast, however, *Arabidopsis* plastocyanin and *Monoraphidium* cytochrome c_6 produce profiles characteristic of other eukaryotes⁶ (Fig. 1c).

In eukaryotic systems, electrostatic interactions between plastocyanin or cytochrome c_6 and photosystem I are dominated by positive charges on the acceptor and negative charges on the donor adjacent to the electron-transfer sites^{2,7}. *Arabidopsis* plasto-

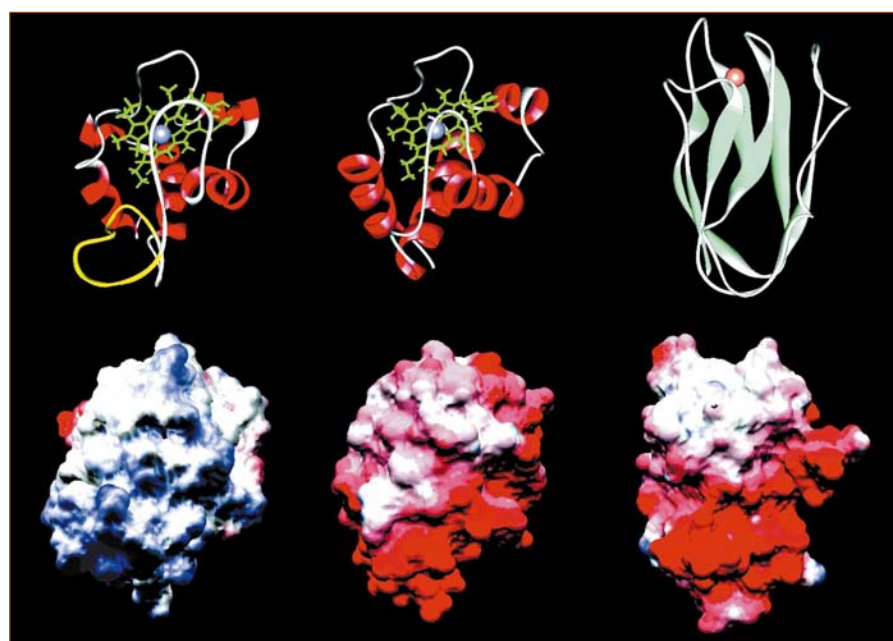


Figure 2 Structural models (top) and surface electrostatic-potential distribution (bottom) of, from left to right, *Arabidopsis* cytochrome c_6 , *Monoraphidium* cytochrome c_6 and *Arabidopsis* plastocyanin. The molecules are similarly orientated: *Monoraphidium* cytochrome c_6 and *Arabidopsis* plastocyanin are placed with their respective negatively charged areas towards the front; *Arabidopsis* cytochrome c_6 is placed with its haem group in the same orientation as that of *Monoraphidium* cytochrome c_6 . Haem groups are shown in green; the extra loop of *Arabidopsis* cytochrome c_6 is shown in yellow; negatively and positively charged regions are shown in red and blue, respectively. Surface electrostatic-potential distributions were calculated at an ionic strength of 40 mM and pH 7.0.

cyanin and *Monoraphidium* cytochrome c_6 both have the negative surface electrostatic potential that typifies eukaryotic proteins, whereas *Arabidopsis* cytochrome c_6 has a positive area close to the solvent-exposed haem (Fig. 2). *Arabidopsis* cytochrome c_6 should therefore be unsuitable as an electron donor to photosystem I, as our results show (Fig. 1).

By contrast, Gupta *et al.*⁴ reported that *Arabidopsis* cytochrome c_6 was as effective as plastocyanin or cytochrome c_6 from *Synechocystis* in stimulating electron transfer.

However, the *Synechocystis* proteins are poor electron donors to higher-plant photosystem I (ref. 6). Although we cannot exclude the possibility that the *Arabidopsis* protein differs from that expressed in *E. coli* (or *Synechocystis*⁴), we know of no modification that could alter the surface electrostatic properties sufficiently to allow reaction with photosystem I. We conclude that *Arabidopsis* cytochrome c_6 is not an effective donor to its own photosystem I. The true function of this molecule may be related to the loop extension of 12 residues (shown in yellow in Fig. 2) that is a unique feature of higher-plant cytochrome c_6 (ref. 3).

Fernando P. Molina-Heredia*, **Jürgen Wastl†**, **José A. Navarro***, **Derek S. Bendall†**, **Manuel Hervás***, **Christopher J. Howe†**, **Miguel A. De la Rosa***

*Instituto de Bioquímica Vegetal y Fotosíntesis, Universidad de Sevilla y Consejo Superior de Investigaciones Científicas, Centro Isla de la Cartuja, Américo Vespucio s/n, 41092-Sevilla, Spain
e-mail: marosa@us.es

†Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, UK

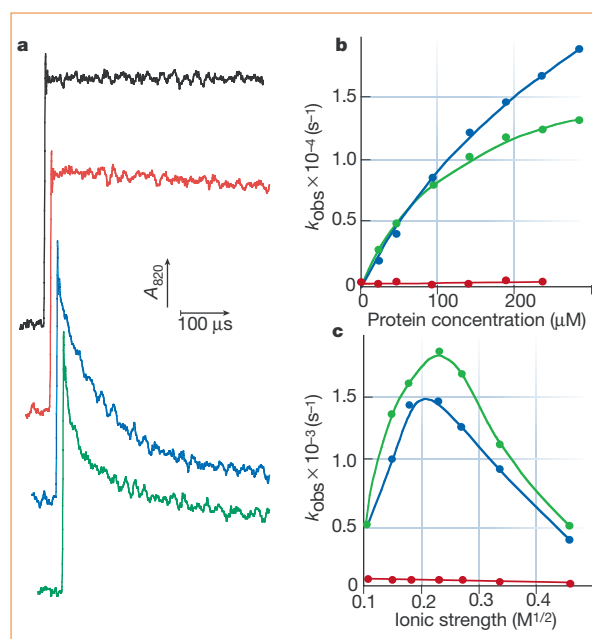


Figure 1 Reduction of *Arabidopsis* photosystem I by cytochrome c_6 of *Arabidopsis* (red) or *Monoraphidium* (blue) and by plastocyanin of *Arabidopsis* (green). **a**, Laser flash-induced kinetic traces measured at 820 nm with 150 μ M protein as electron donor, and a control without added protein (black). **b**, Dependence of the observed rate constant (k_{obs}) on concentration of electron-donor protein. **c**, Effect of ionic strength on k_{obs} with 30 μ M electron-donor protein; ionic strength was adjusted with concentrated NaCl solution. Experiments were carried out at 25 $^{\circ}$ C and pH 7.5. Plant recombinant haem protein was prepared using a sequence encoding mature *Arabidopsis* cytochrome c_6 fused to that of the signal peptide from *Anabaena* cytochrome c_6 , and the synthetic gene was cloned and expressed in *Escherichia coli*; *Monoraphidium* cytochrome c_6 was isolated from cell cultures. Further details are available from the authors.

- Hope, A. B. *Biochim. Biophys. Acta* **1456**, 5–26 (2000).
- De la Rosa, M. A. *et al. Bioelectrochemistry* **55**, 41–45 (2002).
- Wastl, J., Bendall, D. S. & Howe, C. J. *Trends Plant Sci.* **7**, 244–245 (2002).
- Gupta, R., He, Z. & Luan, S. *Nature* **417**, 567–571 (2002).
- Metzger, S. U., Pakrasi, H. B. & Whitmarsh, J. in *Photosynthesis: From Light to Biosphere* Vol. 2 (ed. Mathis, P.) 823–826 (Kluwer Academic, Dordrecht, The Netherlands, 1995).
- Hervás, M., Navarro, J. A., Díaz, A., Bottin, H. & De la Rosa, M. A. *Biochemistry* **34**, 11321–11326 (1995).
- Navarro, J. A., Hervás, M. & De la Rosa, M. A. *J. Biol. Inorg. Chem.* **2**, 11–22 (1997).

Structure of the core domain of human cardiac troponin in the Ca^{2+} -saturated form

Soichi Takeda*†‡, Atsuko Yamashita*, Kayo Maeda* & Yuichiro Maéda*

* Laboratory for Structural Biochemistry, RIKEN Harima Institute at SPring-8, 1-1-1 Kouto, Mikazuki, Sayo, Hyogo 679-5148, Japan

† PRESTO, Japan Science and Technology Corporation (JST), Kawaguchi, Saitama 332-0012, Japan

Troponin is essential in Ca^{2+} regulation of skeletal and cardiac muscle contraction. It consists of three subunits (TnT, TnC and TnI) and, together with tropomyosin, is located on the actin filament. Here we present crystal structures of the core domains (relative molecular mass of 46,000 and 52,000) of human cardiac troponin in the Ca^{2+} -saturated form. Analysis of the four-molecule structures reveals that the core domain is further divided into structurally distinct subdomains that are connected by flexible linkers, making the entire molecule highly flexible. The α -helical coiled-coil formed between TnT and TnI is integrated in a rigid and asymmetric structure (about 80 Å long), the IT arm, which bridges putative tropomyosin-anchoring regions. The structures of the troponin ternary complex imply that Ca^{2+} binding to the regulatory site of TnC removes the carboxy-terminal portion of TnI from actin, thereby altering the mobility and/or flexibility of troponin and tropomyosin on the actin filament.

In muscle contraction, generation of force is coupled with ATPase and is exerted by the interaction between myosin subfragment 1 and the actin-containing muscle thin filament. In skeletal and cardiac muscle, contraction is regulated by intracellular Ca^{2+} concentration. The molecular basis for Ca^{2+} regulation was first established by Ebashi and co-workers in the 1960s, where they identified troponin and tropomyosin as the principal proteins of Ca^{2+} regulation¹. Troponin (relative molecular mass of approximately 80,000 (80K)) and tropomyosin are generally located along polymerized actin, which forms the backbone of the filament, at a troponin:tropomyosin:actin ratio of 1:1:7 (ref. 2). Troponin consists of three subunits: TnC, the Ca^{2+} -binding subunit; TnI, the inhibitory subunit; and TnT, the tropomyosin-binding subunit^{3–5}. Tropomyosin is a α -helical coiled-coil protein throughout its entire length, and interacts with adjacent tropomyosin molecules in a head-to-tail manner, forming continuous strands that lie along the thin filament⁶.

The crystal structure of TnC in complex with the amino-terminal segment of TnI provided insight into the mechanism of calcium regulation⁷. TnC consists of two globular lobes^{8,9} connected by a flexible linker¹⁰. In muscle cells, the C lobe is believed to be continuously occupied with metal ions⁴, and it opens up the hydrophobic patch to bind the N-terminal amphiphilic α -helix of TnI (the first TnC-binding site), which anchors TnC to the rest of troponin, and thus has a structural role^{7,11}. At low Ca^{2+} concentrations, the inhibitory region (residues 137–148 in human cardiac TnI) interacts with actin, and thereby inhibits actomyosin ATPase^{12–14}. On the basis of the structural details of binding in the C lobe, together with the calcium-saturated TnC structure¹⁰, the same manner of interactions in the N lobe was proposed⁷—with an increase in intracellular Ca^{2+} concentration, Ca^{2+} binding to the N lobe would open up the hydrophobic patch to bind to the second amphiphilic α -helix of TnI, which is immediately downstream of the inhibitory region. The inhibitory region is then detached from actin-tropomyosin, resulting in the release of inhibition. The interaction between the N lobe and the TnI amphiphilic segment was confirmed by an NMR study¹⁵.

Although there is little doubt that the changes in the interactions between TnC and TnI have primary roles in the onset of muscle contraction, the mechanism by which troponin–tropomyosin regulates the actomyosin ATPase has remained elusive. This is because high-resolution structures have been available for only a small portion of the troponin–tropomyosin–actin complex. Electron microscopy^{16–18} and low-resolution X-ray crystallography¹⁹ studies have revealed that troponin consists of two domains: the TnT1 extension and the rest of troponin, referred to as the core domain. We have chosen the core domain for the target of crystallographic study, as this retains most of the regulatory function of troponin²⁰. Here we describe the crystal structure of two kinds of core domain of human cardiac troponin in the Ca^{2+} -saturated form; one at 2.6 Å resolution and the other at 3.3 Å resolution. These crystal structures elucidate how three polypeptide chains are folded around each other within the troponin molecule.

Structure determination

We crystallized two distinct preparations of the troponin core domain, which were reconstituted from *Escherichia coli*-expressed recombinant human cardiac troponin subunits. One core domain had a relative molecular mass of 46K (Tn46K), consisting of TnC(1–161), TnI(31–163) and TnT(183–288), whereas the other was 52K (Tn52K), consisting of TnC(1–161), TnI(31–210) and TnT(183–288). Our crystallization trials of the entire troponin molecule were not successful, presumably due to the association of the extended structure of TnT1, which must have an extended binding site to tropomyosin^{17–19,21}. Both crystal forms grew in the presence of Ca^{2+} , so that all of the metal-binding sites were occupied (Ca^{2+} -bound form). In both cases, the cardiac-specific sequence (TnI residues 1–30) was removed, which makes each subunit almost identical to the skeletal muscle troponin counterpart. In Tn46K, but not in Tn52K, an extra 46 residues were deleted from the C terminus of TnI, because the deletion improved the crystal quality. However, this deletion is known to impair the ability of TnI to bind to actin–tropomyosin at low Ca^{2+} concentrations²⁰.

The Tn46K structure was determined by the multi-wavelength anomalous dispersion (MAD) method, and was refined against the native crystal data up to 2.6 Å to an *R*-factor of 25.9% (*R*_{free} 29.7%). The Tn52K structure was determined by molecular replacement

‡ Present address: Department of Cardiac Physiology, National Cardiovascular Center Research Institute, 5-7-1 Fujishiro-dai, Suita, Osaka 565-8565, Japan.

using the Tn46K structure and was refined up to 3.3 Å resolution to an *R*-factor of 25.1% (*R*_{free} 30.8%) (Table 1). The asymmetric unit of each crystal form contained two troponin molecules, and thus we obtained the structures of four molecules in total, which we term Tn46KA, Tn46KB, Tn52KA and Tn52KB.

Overall architecture and subunit arrangement

The overall architecture of the Tn52KB molecule is shown in Fig. 1a. Notably, the core domain is dominated by α-helices. The core domain is further divided into two structurally distinct subdomains (Fig. 1b), denoted as the regulatory head (consisting of TnC residues 3–84 and TnI residues 150–159) and the IT arm (consisting of TnC residues 93–161, TnI residues 42–136 and TnT residues 203–271). All four molecules are similar to each other, except for the relative orientation between the two subdomains (see below).

Each lobe of TnC superposes almost exactly onto its counterpart in the previously reported TnC–TnI binary complexes: Protein Data Bank code 1MXL¹⁵ (the root mean square (r.m.s.) deviation between the equivalent Cα atoms in the TnC N lobe is 1.50 Å) and 1A2X⁷ (the r.m.s. deviation in the TnC C lobe is 1.04 Å).

The N terminus of TnI (residues 31–34) was not well defined in either molecule, and the following residues 35–42 are highly variable among the four molecules. Residues 43–79 form the α-helix, H1(I), and its amphiphilic portion (residues 43–65) binds to the C lobe of TnC through multiple polar and van der Waals interactions (the first TnC-binding site in TnI), as already shown in the TnC–TnI(1–47) binary complex⁷. The C-terminal portion of H1(I) (residues 66–79), which was disordered in the binary complex crystal⁷, is stabilized here by interacting with TnT through multiple hydrogen bonds as well as hydrophobic interactions (Fig. 1c). TnI residues 80–89, which are less conserved and contain two glycines and one proline, have an extended structure and interact with TnT, forming a hydrophobic core (Fig. 1c; Leu 83, Leu 85, Leu 88, Leu 93 and Leu 96 of TnI, and Ala 233, Arg 230 and Leu 229 of TnT are involved in the interaction).

The H2(I) helix (TnI residues 90–135) forms a parallel α-helical coiled-coil with the H2(T2) helix (residues 226–271) of TnT (Fig. 1), as previously suggested^{22,23}. Each chain has 6.5 heptad repeats, with hydrophobic residues at the a and d positions (see Supplementary Fig. 1). The amino acid residues involved in the inter-chain interactions are well conserved even in invertebrate species (see Supplementary Fig. 1), suggesting that this coiled-coil between TnT and TnI has important physiological roles that are characteristic of troponin. At the C terminus of the coiled-coil, TnT (residues 256–270) interacts with TnC (see below).

Downstream from the coiled-coil is the TnI inhibitory region (residues 137–148). In the present structures, the electron densities associated with the inhibitory region were not well defined in either molecule. However, the inhibitory region must be extended judging

from the Cα distances (26.2–32.6 Å, depending on the four molecules) between Arg 136 and Thr 149, spanning 13 residues. The inhibitory region is followed by another amphiphilic α-helix, H3(I), which spans residues 150–159 of TnI, and interacts with the hydrophobic cleft of the N lobe of TnC through multiple van der Waals contacts (the second TnC-binding site in TnI; see below). Further downstream of H3(I), residues 164–188 form a protruded α-helix, H4(I), that has no direct interaction with the rest of the molecule (Fig. 1a, b). H4(I) was clearly defined in only one molecule (Tn52KB) within an asymmetric unit, and was probably stabilized through eventual molecular contacts with the loop between H1(I) and H2(I) of the symmetry-related molecule. On the other hand, Tn52KA lacks a defined protruded α-helix, but has a loop (residues 160–162) pointing in a different direction than that of Tn52KB (Fig. 2a), owing to a kink at Gly 160. Although we found further extended electron densities of TnI in the Tn52KA molecule in this direction, they were quite ambiguous, so we did not include these residues in the model. TnI residues 189–191 of Tn52KB are in an extended structure that eventually interacts with the next molecule, whereas the electron density associated with the C terminus (residues 192–210) was not defined. The C terminus of TnT (denoted as C-TnT, residues 272–288), which has been shown to bind to tropomyosin^{24–26}, is not always defined in the crystal (the residues down to 276 are defined in Tn52KB). Upstream from the coiled-coil, residues 204–220 of TnT form another α-helix, H1(T2), which is kinked by about 60° relative to H2(T2) (Fig. 1a, c). H1(T2) is predominantly stabilized against the rest of the molecule through two sets of hydrogen bonds: one is between Glu 214 of TnT and Arg 98 of TnI, and the other is between Arg 216 of TnT and Asp 105 of TnI in Tn46KA (Fig. 1c) and in Tn52KA (not shown). However, Tn46KB and Tn52KB lack these hydrogen bonds, probably due to the distinct molecular packing in the crystal lattice (data not shown). Further upstream of H1(T2), residues 183–202 of TnT were not defined in the electron density map.

As described above, most of the electron densities associated with the potential actin–tropomyosin interfaces are ambiguous. These include the inhibitory region (residues 137–148) and the C-terminus of TnI (residues 163–210 in Tn52KA and 192–210 in Tn52KB), and the C-TnT (residues 272–288). In total, 12.5% of the Tn46K and 19.4% of the Tn52K amino acid residues are not included in the current models. These regions lack direct interactions with the remainder of the molecule and are susceptible to proteases^{20,24}. It is therefore plausible that these actin–tropomyosin-binding sites may not properly fold in solution, or alternatively, that these regions may contain flexible linkers that allow multiple conformations. The structures from the two crystal forms show relatively high overall *B*-factors (the averaged *B*-factors are 73.0 Å² for Tn46K and 91.7 Å² for Tn52K), consistent with the rapid decrease of the diffraction intensities in the higher angle region. The high *B*-factors are probably derived from the extremely high flexibility of the molecule at the subdomain junction (see below). The high overall *B*-factors, together with the substantial extension of the disordered regions, might contribute to the relatively high *R*-factors. However, the experimentally phased electron density maps (data not shown) and the final 2*F*_o–*F*_c map (Fig. 1d, e) seemed to be of sufficient quality for the structural determination. The unusual flexibility is a remarkable characteristic of the troponin molecule. As troponin is quite flexible, the molecule is able to evade packing in a crystalline lattice for a long time.

Flexible linkers between the two subdomains

In Fig. 2, the four molecules are superposed, with a fixed orientation of either the regulatory head (Fig. 2a) or the IT arm (Fig. 2b). The r.m.s. deviations between all of the equivalent Cα atoms of any two of the four troponin molecules are 0.52–0.81 Å for the regulatory head and 0.68–1.70 Å for the IT arm. Except for the loop (TnI residues 81–89) in Tn52KB, the atomic coordinates within each

Table 1 Refinement statistics

	Core domains	
	Tn46K	Tn52K
Resolution (Å)	40.0–2.61	20.0–3.30
Reflections	27,881	16,015
<i>R</i> ^a	0.264	0.251
<i>R</i> _{free} ^a	0.298	0.308
Number of atoms		
Protein	5,673	5,858
Water	102	0
Calcium ion	6	6
Average <i>B</i> (Å ²)	73.0	91.7
Ramachandran plot		
Most favoured/allowed (%)	91.8/8.2	91.5/8.5
r.m.s. deviations		
Bond length (Å)/bond angle (°)	0.0085/1.398	0.0089/1.40

^a *R* = Σ||*F*_o| – |*F*_c||/Σ|*F*_o|. *R*_{free} is the *R* value for a subset of 5% of the reflection data, which were not included in the crystallographic refinement.

subdomain are almost identical among the four molecules, indicating that each subdomain behaves as a unit. On the other hand, the relative orientation between the two subdomains is quite variable among the four molecules. The largest difference exists between Tn46KA and Tn46KB, where the IT arm should be rotated by 20° relative to the regulatory head subdomain, bringing about a 27 Å displacement at the distal end of the IT arm (Fig. 2a).

On the basis of this observation and the following considerations,

we suggest that the linker between the two subdomains works as a universal joint. First, the unique covalent link between the two subdomains—that is, the linker connecting the two lobes of TnC (D/E-linker)—must be highly mobile. The relative orientation between the two lobes in our structures is completely different from those reported previously^{7–9,27} (Fig. 3). In our study, the D/E-linker in each molecule is not clearly defined owing to the lack of specific interactions with the rest of the molecule. TnC mutants

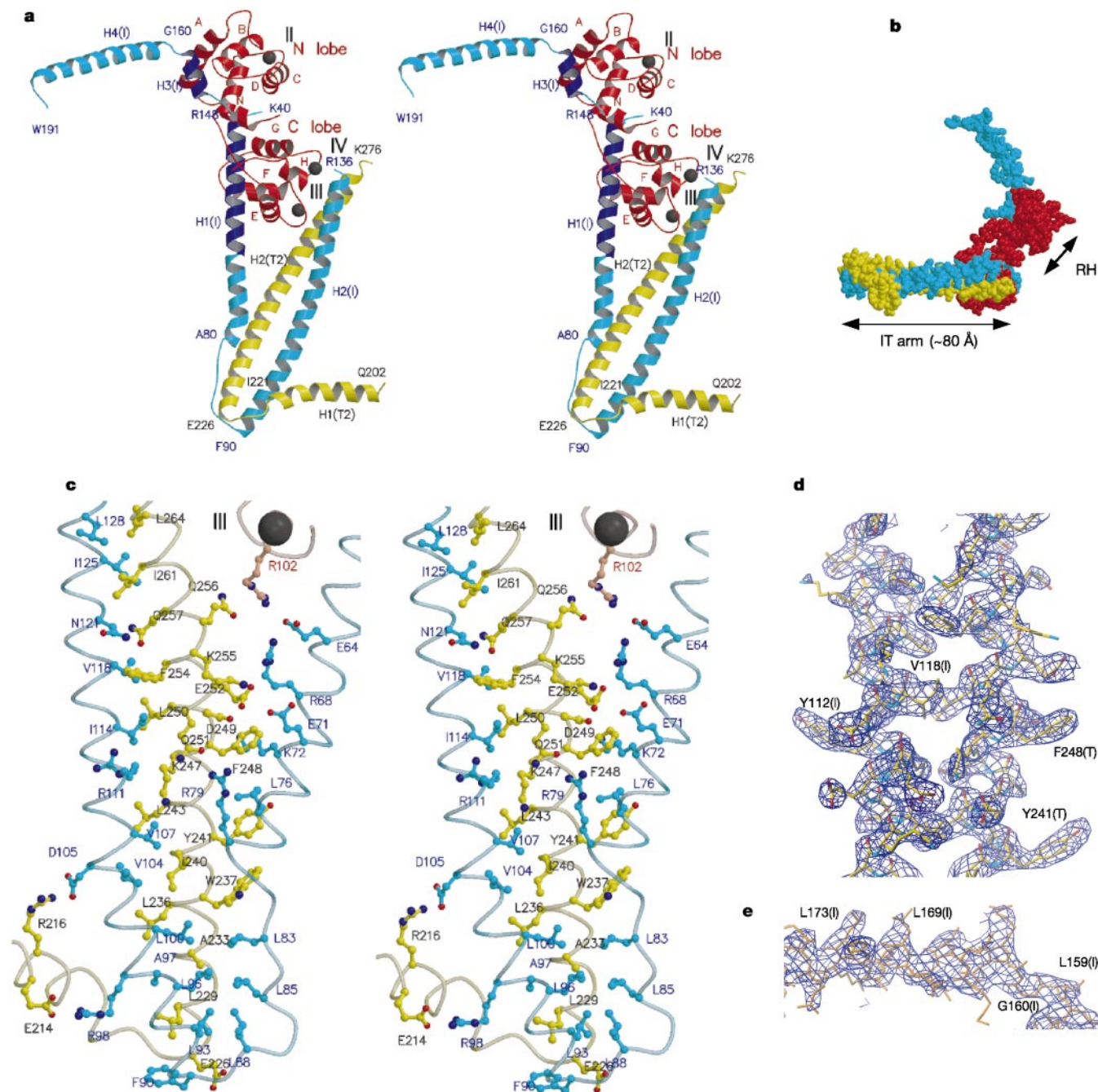


Figure 1 Crystal structure of the troponin core domain. **a**, Stereo view of the Tn52KB molecule. TnC and TnT are coloured in red and yellow, respectively. TnI is coloured in cyan, except for the two stretches of amphiphilic helices (TnC-binding sites), which are dark blue. The three Ca^{2+} ions bound to the Ca^{2+} -binding sites (II–IV) are represented by black spheres. Each helix within TnI and TnT is indicated by the helix number, whereas each helix of TnC is indicated by a capital letter (N and A–H). **b**, A space-filling model of

the Tn52KB molecule. RH, regulatory head. **c**, A close-up stereo view of the IT-arm portion (Tn46KA), viewed from the other side of **a**. **d**, $2F_o - F_c$ electron density map (1.5σ contoured) associated with the middle region of the coiled-coil at 2.6 Å resolution (Tn46KA). **e**, $2F_o - F_c$ map (1.2σ contoured) associated with the protruded helix, H4(I), at 3.3 Å resolution (Tn52KB). Figures 1–5 were generated by TURBO-FRODO⁴⁶, MOLSCRIPT⁴⁹ and Raster3D⁵⁰.

with D/E-linker deletions do not show a significant impairment in regulatory activity, whereas mutants that reduce the flexibility of the D/E-linker impair the ability to activate the thin filament^{28,29}. Thus, the flexibility, not the length (which would specify the relative orientation between the two lobes), of the D/E-linker is crucial for the activity of troponin. Second, the direct contacts between the regulatory head and the IT arm are mediated by a small number of residues (data not shown) that are poorly conserved among species, and therefore the contact may not be specific, but would eventually be stabilized by the crystal-packing force. Third, the link through TnI, the inhibitory region that connects H2(I) and H3(I), is not well defined in the present structures, indicating that this link is also flexible. Altogether, these findings indicate that the troponin molecule might adopt multiple conformations with variable subdomain orientations when it is located on the filament, as well as within the crystal.

Integration of the C lobe of TnC

At the C terminus of the coiled-coil, TnT interacts with the C lobe of TnC (Fig. 4a). Therefore, all three polypeptide chains of troponin come together in this region to interact in a specific manner. Here, helix H2(T2) interacts with TnI by forming a coiled-coil on one side. The other side of H2(T2) interacts with the Ca²⁺-binding loops of TnC: TnT residues Tyr 259, Val 263, Asn 266, Arg 267 and Asp 270 interact with TnC residues Phe 101, Arg 102, Asp 105, Asp 109 and Tyr 111 within loop III, and with Asp 149, Tyr 150 and Asp 151 within loop IV. Notably, Asp 270 of TnT forms a hydrogen bond with the hydroxyl group of Tyr 111 of TnC, which simultaneously coordinates the calcium ion in site III with its carbonyl oxygen. Moreover, although the side chain of Asp 109 is also involved in calcium binding, its carbonyl oxygen also forms a hydrogen bond with Asn 266 of TnT. These findings suggest that the binding of Ca²⁺ ions to the C lobe is required not only to keep the C lobe tightly folded to bind to the amphiphilic helix of TnI (the first TnC-binding site in TnI), but also to stabilize the interaction between TnC and TnT. As the metal-binding sites in the C lobe (site III and IV) are believed to be occupied with either Ca²⁺ or Mg²⁺ ions *in vivo*, irrespective of the sarcoplasmic Ca²⁺ concentration⁴, the C lobe is rigidly integrated in the IT arm, and this structure remains unchanged and is independent of the physiological states of the thin filament. It is worth noting that the integration site of three chains resides immediately upstream of the C-terminal region of TnT (C-TnT, residues 272–288) through which the structure is anchored to actin–tropomyosin^{24–26}, and the C-terminal region of TnI (TnI_{reg},

residues 137–210), which must undergo major conformational changes depending on the Ca²⁺ concentration (see below).

Molecular switch

There is little doubt that the amphiphilic helix H3(I) works as a molecular switch that transmits the initial signal of Ca²⁺ binding to the N lobe of TnC to the other components in the thin filament, and thus has a central role in the initial steps of troponin–tropomyosin-based regulation. As discussed previously⁷, amphiphilic helix H3(I) binds specifically to the hydrophobic patch of the Ca²⁺-saturated N lobe of TnC (Fig. 4b) through conserved hydrophobic residues (Supplementary Fig. 1b), and the interactions are Ca²⁺-dependent^{15,20}. The segment H3(I) is located in the sequence between the two putative actin-binding sites, the inhibitory region (residues 137–148) and the C terminus of TnI (residues 169–210)²⁰, which were both shown to be essential for the inhibitory binding of TnI in the absence of Ca²⁺ (ref. 14) and to change the distances to the nearby actin molecule depending on the Ca²⁺ concentrations^{30,31}. Residues 167–184 are highly conserved (Supplementary Fig. 1b) and form the amphiphilic helix, H4(I); the corresponding peptide (residues 128–148 in skeletal TnI) binds to actin–tropomyosin and induces a weak inhibitory activity on its own³². Finally, in the presence of Ca²⁺, TnI is released from its binding site on actin–tropomyosin³³. Therefore, the binding of H3(I) to the hydrophobic patch of the N lobe of TnC must induce the detachment of the substantially extended C-terminal portion of TnI, denoted as the regulatory segment of TnI, TnI_{reg} (residues 137–210), from the actin filament.

Interactions of troponin and other thin filament components

Troponin is anchored to the thin filament mainly through tropomyosin binding of two distinct portions of TnT; that is, TnT1 and C-TnT^{16,24–26}. The former is believed to be the major Ca²⁺-insensitive anchoring site of troponin onto tropomyosin, whereas the interaction through the latter is stronger at low, rather than high, Ca²⁺ concentrations²⁵ (Fig. 5). No part of the IT arm is known to have direct interactions with actin or with tropomyosin. On the basis of the present structures, the separation between TnT1 and C-TnT is estimated to be about 60 Å (Fig. 5), although these two portions are not included in the current crystal structures. TnT1 has a high α -helical content and is less susceptible to proteolysis, indicating that it forms another structural domain^{16,19}. TnT residues 183–200 probably work as a flexible linker between the two subdomains, TnT1 and the IT arm. This segment is not defined in

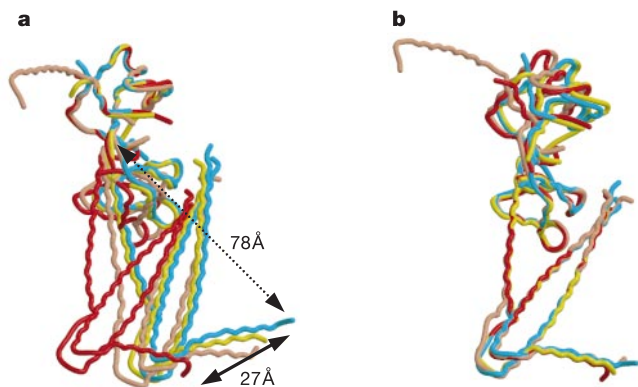


Figure 2 Comparison of the four troponin core domain molecules from two crystal forms, Tn46K and Tn52K. The four molecules are superposed, with the best superposition (the least square fit) of either the regulatory head (a) or the IT arm (b). The molecules Tn46KA, Tn46KB, Tn52KA and Tn52KB are coloured red, cyan, yellow, and light pink, respectively.

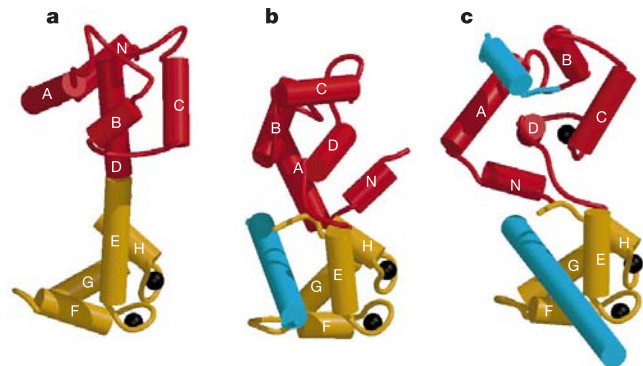


Figure 3 Comparison of the TnC conformations, represented by cylinders (α -helices) and wires (loops). a, Crystal structure of TnC alone (Protein Data Bank code 1TOP9); b, TnC in complex with the N-terminal fragment of TnI (1A2X7); and c, TnC in the present study (Tn46KA). The C lobes are aligned in almost the same orientation. The N and C lobes are red and orange, respectively, whereas the bound TnI segments are cyan. Bound calcium ions are represented by black spheres. The TnC helices, N and A–H, are indicated.

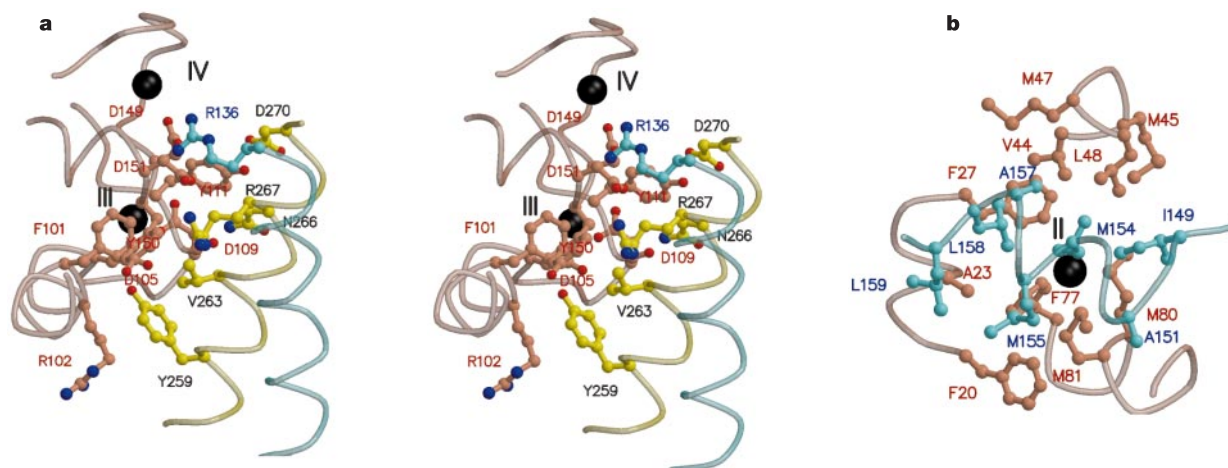


Figure 4 Side-chain interactions between the subunits. **a**, A stereo view of the interactions between the three chains around the C lobe of TnC. **b**, The hydrophobic residues in the N lobe of TnC interacting with segment H3(I) of TnI. TnC, TnI and TnT are

coloured pink, cyan and yellow, respectively. The amino acid residues involved in the interactions are shown in ball and stick representation. The three Ca^{2+} ions bound to sites II, III and IV are shown by the black spheres.

the present structures, is less conserved (Supplementary Fig. 1a) and is susceptible to proteolysis^{16,26}.

At a low Ca^{2+} concentration, the N lobe of TnC dissociates H3(I). It has been interpreted that TnI should form another attachment to the filament, most probably to create the ternary complex actin–tropomyosin–TnI^{32,33}, whereas others have shown that TnI cross-links only to actin³⁰.

Discussion

The present structures have elucidated the overall architecture of the troponin molecule. Troponin appears to be divided into subdomains, including the regulatory head, the IT arm, TnT1, C-TnT and TnI_{reg}. The subdomains are defined by boundaries that do not coincide at all with the boundaries of the three polypeptide chains. The subdomains are connected by flexible linkers, making the entire molecule quite flexible. The flexible nature of the troponin molecule must be relevant to the physiological function. Although many

residues are left unidentified in the present electron density maps and some residues are quite variable among the four molecules, the atomic positions in the remaining parts are reliable, partly because the structure has been solved four times. Actually, the missing residues and the differentiated relative orientations of individual subdomains among four molecules indicate that the linkers are flexible. The characteristic architecture implies that molecular motions of troponin could be described in terms of changes of orientation of individual α -helices plus mobility of individual flexible linkers.

During calcium regulation, TnI_{reg} must undergo major changes both in position and conformation. At higher Ca^{2+} concentrations, TnI_{reg} is detached from actin–tropomyosin, being associated with the N lobe of TnC in an extended form as shown in the present structures. At lower Ca^{2+} concentrations, TnI_{reg} must form an extra attachment to actin–tropomyosin so that the troponin–tropomyosin strand is tied down onto the actin filament. Within the extra attachment, the entire TnI_{reg} might be folded, as complete inhibition requires an extended region, at least residues spanning 185–210 (refs 14, 20), and as the space available on actin–tropomyosin might be limited if TnI_{reg} binds to one of the seven actin monomers.

The IT arm must also have an important function in calcium regulation. This subdomain is large and rigid, being conserved between species and having no direct interaction with actin–tropomyosin. Its location is remarkable—this structure bridges the two (largely) Ca^{2+} -independent attachments to actin–tropomyosin, TnT1 and C-TnT. Moreover, the IT arm resides immediately upstream, not only from the tropomyosin-binding site (C-TnT) but also from the mobile TnI_{reg} that changes position and conformation in a Ca^{2+} -dependent manner, promoting the transition from two points of attachment to three points of attachment of troponin to actin–tropomyosin. An intriguing possibility is that the formation of the third attachment (by TnI_{reg}) could cause a minute rotation of the IT arm about the pivotal point; that is, the C-terminal end of the coiled-coil. The formation of the third attachment itself, as well as the rotation of the IT arm, might change the properties of the tropomyosin strand on the actin filament.

The nature of the changes remain unknown. Although many previous results have been interpreted to be consistent with the proposal^{34–36} for the shift in azimuthal position of the tropomyosin strands, the systematic FRET experiments³⁷ consistently failed to provide evidence of any major change in the distance between tropomyosin and actin. A more plausible explanation is that the strain imposed on the tropomyosin strand is altered; thereby the

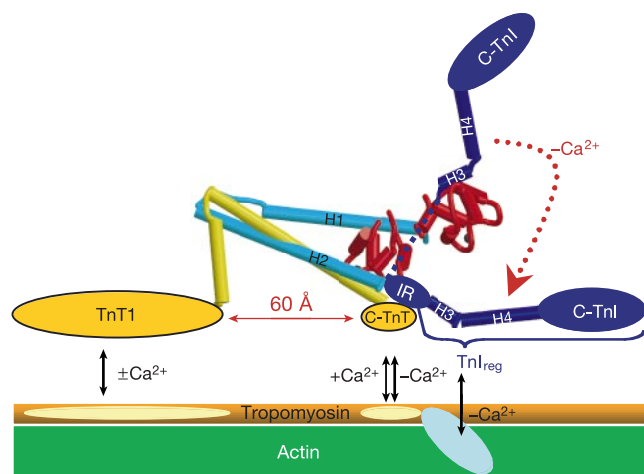


Figure 5 A schematic representation of the interactions between troponin and other thin filament components. The potential actin–tropomyosin-binding portions, which are not included in the current structural model, are schematically drawn: TnT1 and C-TnT are yellow ellipsoids, and the inhibitory region (IR) and the C terminus of TnI (C-TnI) are blue ellipsoids. The actin and the tropomyosin strands are in green and brown, respectively. The black arrows indicate the interactions between troponin and tropomyosin–actin.

mobility and/or the flexibility of the tropomyosin strand might be changed³⁸. Being consistent with this view, the atomic model of the actin–tropomyosin complex (without troponin)³⁹ indicated that the tropomyosin strand ‘diffuses’ around actin, being weakly held by electrostatic forces. Further investigations are required to elucidate the structure of the troponin–tropomyosin complex corresponding to each state in the three-state model⁴⁰. In this context, it is worth noting that the thin-filament-associated X-ray diffraction intensities as well as electron micrographs have been interpreted based on the assumption that the mass of the troponin–tropomyosin complex is distributed evenly and smoothly all through the continuous ‘tropomyosin strands’. With the present atomic structures, trials are now underway to identify separately the mass of tropomyosin and that of troponin on the muscle thin filament. □

Methods

Protein preparations

Details about the construction of the expression vectors and the sample preparation will be described elsewhere. We prepared individual subunits of human cardiac troponin separately in *E. coli*. TnT2 (residues 183–288) was prepared by chemical cleavage. First, TnT(25K)⁴¹ (TnT residues 94–288) was expressed and purified. Then TnT(25K) was further cleaved by cyanogen bromide to obtain TnT2, which was purified by reverse phase high-performance liquid chromatography (HPLC) followed by size exclusion chromatography. As previously demonstrated for the TnI(1–47) fragment⁷, when preparing peptide fragments, the chemical cleavage method gives rise to more homogenous and stable preparations that lack proteolytic activity as compared with limited proteolysis or direct expression of the fragment in *E. coli*. For TnC, the cysteine-less variant TnC(C35S/C84S) was used. For TnI, the variants (T31M/C80A/C97A) were used, which each start from Met 31. The ternary complexes were prepared²⁰ with minor modifications. For preparing the strontium-substituted Tn46K crystals, SrCl₂ instead of CaCl₂ was added to the solutions throughout the purification procedure. The final protein solutions contained 10 mM Tris-HCl, 150 mM NaCl and either 0.1 mM CaCl₂ or 0.1 mM SrCl₂ at pH 8.

Crystallization

Native and strontium-substituted crystals were grown using the hanging-drop vapour diffusion technique, with reservoir solutions containing 20% PEG3350, 15% glycerol, 0.1 M LiCl, 50 mM Tris-HCl and either 5 mM CaCl₂ or 5 mM SrCl₂ at pH 8. Drops composed of 2 µl of protein at 10 mg ml⁻¹ and 2 µl of the reservoir solution were equilibrated with 0.5 ml of reservoir solution for a minimum of two weeks at 20 °C. Tn46K and Tn52K crystallized in space group P2₁ with unit cell dimensions of $a = 42.3 \text{ \AA}$, $b = 167.9 \text{ \AA}$, $c = 69.7 \text{ \AA}$, $\beta = 101.4^\circ$ and $a = 48.3 \text{ \AA}$, $b = 169.5 \text{ \AA}$, $c = 68.5 \text{ \AA}$, $\beta = 102.4^\circ$, respectively. The osmium-derivative crystal was prepared by soaking the native Tn46K crystal in a reservoir solution supplemented with 1 mM OsCl₃ for 18 h. The crystals were flash frozen under a nitrogen flow at 100 K after the glycerol concentration was increased to 20% by soaking.

Diffraction data collection and structural analysis

All diffraction data were collected at Spring-8 (beam lines BL41XU, BL44B2 and BL45XU) using either Mar CCD (charge-coupled device) 165 or Rigaku R-AXIS V detectors at 90 K. The images were reduced using HKL2000 (ref. 42). The initial phases were obtained at 3.3 Å resolution by the MAD method, using an osmium derivative data set. The heavy atom positions were searched for by SOLVE⁴³, and the phases were calculated using SHARP⁴⁴ and further improved by SOLMON⁴⁵. Model building was carried out using TURBO-FRODO⁴⁶. Although the Os MAD map enabled us to trace almost the entirety of the chains, the quality of the map was not sufficient to complete the model building in detail. Then, we obtained new phases to 2.8 Å resolution using a set of MAD data from a strontium-substituted crystal, which enabled us to complete the model building. In the crystal, six Sr²⁺ ions were coordinated in the Ca²⁺-binding sites (II–IV). The structural models (Protein Data Bank codes 1MXL¹⁵ and 1A2X⁷) were used as guides for the model building. The model was refined against the native Tn46K data set to 2.6 Å resolution using CNS⁴⁷ (including the bulk solvent correction, the simulated annealing, the minimization and the individual B-factor refinements), and the subsequent rounds of model building and refinement produced the final structural model for two Tn46K molecules.

The structure of Tn52K was solved by molecular replacement, using the refined Tn46K model. To do this, the Tn46K subdomains, the IT arm and regulatory head were separately used as the search models by the program MOLREP in the CCP4 suite⁴⁸. The subsequent rounds of model building and refinement (as described above plus the grouped-B-factor refinement) produced the final structural model for the two Tn52K molecules.

Received 24 December 2002; accepted 28 April 2003; doi:10.1038/nature01780.

1. Ebashi, S. & Endo, M. Calcium ion and muscle contraction. *Prog. Biophys. Mol. Biol.* **18**, 123–183 (1968).
2. Ebashi, S., Endo, M. & Ohtsuki, I. Control of muscle contraction. *Q. Rev. Biophys.* **2**, 351–384 (1969).
3. Ohtsuki, I., Maruyama, K. & Ebashi, S. Regulatory and cytoskeletal proteins of vertebrate skeletal muscle. *Adv. Protein Chem.* **38**, 1–67 (1986).

4. Zot, A. S. & Potter, J. D. Structural aspects of troponin–tropomyosin regulation of skeletal muscle contraction. *Annu. Rev. Biophys. Biophys. Chem.* **16**, 535–559 (1987).
5. Farah, C. S. & Reinach, F. C. The troponin complex and regulation of muscle contraction. *FASEB J.* **9**, 755–767 (1995).
6. Phillips, G. N. Jr, Fillers, J. P. & Cohen, C. Tropomyosin crystal structure and muscle regulation. *J. Mol. Biol.* **192**, 111–131 (1986).
7. Vassilyev, D. G., Takeda, S., Wakatsuki, S., Maeda, K. & Maeda, Y. Crystal structure of troponin C in complex with troponin I fragment at 2.3-Å resolution. *Proc. Natl Acad. Sci. USA* **95**, 4847–4852 (1998).
8. Herzberg, O. & James, M. N. Structure of the calcium regulatory muscle protein troponin-C at 2.8 Å resolution. *Nature* **313**, 653–659 (1985).
9. Sundaralingam, M. *et al.* Molecular structure of troponin C from chicken skeletal muscle at 3-angstrom resolution. *Science* **227**, 945–948 (1985).
10. Slupsky, C. M. & Sykes, B. D. NMR solution structure of calcium-saturated skeletal muscle troponin C. *Biochemistry* **34**, 15953–15964 (1995).
11. Gasmí-Seabrook, G. M. *et al.* Solution structures of the C-terminal domain of cardiac troponin C free and bound to the N-terminal domain of cardiac troponin I. *Biochemistry* **38**, 8313–8322 (1999).
12. Syska, H., Wilkison, J. M., Grand, R. J. & Perry, S. V. The relationship between biological activity and primary structure of troponin I from white skeletal muscle of the rabbit. *Biochem. J.* **153**, 375–387 (1976).
13. Talbot, J. A. & Hodges, R. S. Synthetic studies on the inhibitory region of rabbit skeletal troponin I. Relationship of amino acid sequence to biological activity. *J. Biol. Chem.* **256**, 2798–2802 (1981).
14. Farah, C. S. *et al.* Structural and regulatory functions of the NH₂- and COOH-terminal regions of skeletal muscle troponin I. *J. Biol. Chem.* **269**, 5230–5240 (1994).
15. Li, M. X., Spyropoulos, L. & Sykes, B. D. Binding of cardiac troponin-I147–163 induces a structural opening in human cardiac troponin-C. *Biochemistry* **38**, 8289–8298 (1999).
16. Ohtsuki, I. Molecular arrangement of troponin-T in the thin filament. *J. Biochem. (Tokyo)* **86**, 491–497 (1979).
17. Flicker, P. E., Phillips, G. N. Jr & Cohen, C. Troponin and its interactions with tropomyosin. An electron microscope study. *J. Mol. Biol.* **162**, 495–501 (1982).
18. Ohtsuki, I., Onoyama, Y. & Shiraiishi, F. Electron microscopic study of troponin. *J. Biochem. (Tokyo)* **103**, 913–919 (1988).
19. White, S. P., Cohen, C. & Phillips, G. N. Jr Structure of co-crystals of tropomyosin and troponin. *Nature* **325**, 826–828 (1987).
20. Takeda, S., Kobayashi, T., Taniguchi, H., Hayashi, H. & Maeda, Y. Structural and functional domains of the troponin complex revealed by limited digestion. *Eur. J. Biochem.* **246**, 611–617 (1997).
21. Schaertl, S., Lehrer, S. S. & Geeves, M. A. Separation and characterization of the two functional regions of troponin involved in muscle thin filament regulation. *Biochemistry* **34**, 15890–15894 (1995).
22. Pearlstone, J. R. & Smillie, L. B. The interaction of rabbit skeletal muscle troponin-T fragments with troponin-I. *Can. J. Biochem. Cell Biol.* **63**, 212–218 (1985).
23. Stefancsik, R., Jha, P. K. & Sarkar, S. Identification and mutagenesis of a highly conserved domain in troponin T responsible for troponin I binding: potential role for coiled coil interaction. *Proc. Natl Acad. Sci. USA* **95**, 957–962 (1998).
24. Tanokura, M., Tawada, Y., Ono, A. & Ohtsuki, I. Chymotryptic subfragments of troponin T from rabbit skeletal muscle. Interaction with tropomyosin, troponin I and troponin. *C. J. Biochem. (Tokyo)* **93**, 331–337 (1983).
25. Pearlstone, J. R. & Smillie, L. B. Effects of troponin-I plus-C on the binding of troponin-T and its fragments to alpha-tropomyosin. Ca²⁺ sensitivity and cooperativity. *J. Biol. Chem.* **258**, 2534–2542 (1983).
26. Morris, E. P. & Lehrer, S. S. Troponin-tropomyosin interactions. Fluorescence studies of the binding of troponin, troponin T, and chymotryptic troponin T fragments to specifically labeled tropomyosin. *Biochemistry* **23**, 2214–2220 (1984).
27. Houdusse, A., Love, M. L., Dominguez, R., Grabarek, Z. & Cohen, C. Structures of four Ca²⁺-bound troponin C at 2.0 Å resolution: further insights into the Ca²⁺-switch in the calmodulin superfamily. *Structure* **5**, 1695–1711 (1997).
28. Ramakrishnan, S. & Hitchcock-DeGregori, S. E. Investigation of the structural requirements of the troponin C central helix for function. *Biochemistry* **34**, 16789–16796 (1995).
29. Babu, A., Rao, V. G., Su, H. & Gulati, J. Critical minimum length of the central helix in troponin C for the Ca²⁺ switch in muscular contraction. *J. Biol. Chem.* **268**, 19232–19238 (1993).
30. Luo, Y. *et al.* Photocrosslinking of benzophenone-labeled single cysteine troponin I mutants to other thin filament proteins. *J. Mol. Biol.* **296**, 899–910 (2000).
31. Li, Z., Gergely, J. & Tao, T. Proximity relationships between residue 117 of rabbit skeletal troponin-I and residues in troponin-C and actin. *Biophys. J.* **81**, 321–333 (2001).
32. Tripet, B., Van Eyk, J. E. & Hodges, R. S. Mapping of a second actin-tropomyosin and a second troponin C binding site within the C terminus of troponin I, and their importance in the Ca²⁺-dependent regulation of muscle contraction. *J. Mol. Biol.* **271**, 728–750 (1997).
33. Geeves, M. A., Chai, M. & Lehrer, S. S. Inhibition of actin-myosin subfragment 1 ATPase activity by troponin I and IC: relationship to the thin filament states of muscle. *Biochemistry* **39**, 9345–9350 (2000).
34. Huxley, H. E. Structural changes in the actin- and myosin-containing filaments during contraction. *Cold Spring Harbor Symp. Quant. Biol.* **37**, 361–376 (1972).
35. Haselgrove, J. C. X-ray evidence for a conformational change in the actin-containing filaments of vertebrate striated muscle. *Cold Spring Harbor Symp. Quant. Biol.* **37**, 341–352 (1972).
36. Parry, D. A. & Squire, J. M. Structural role of tropomyosin in muscle regulation: analysis of the x-ray diffraction patterns from relaxed and contracting muscles. *J. Mol. Biol.* **75**, 33–55 (1973).
37. Hai, H., Sano, K., Maeda, K., Maeda, Y. & Miki, M. Ca²⁺- and S1-induced conformational changes of reconstituted skeletal muscle thin filaments observed by fluorescence energy transfer spectroscopy: structural evidence for three states of thin filament. *J. Biochem. (Tokyo)* **131**, 407–418 (2002).
38. Lehrer, S. S., Golitsina, N. L. & Geeves, M. A. Actin-tropomyosin activation of myosin subfragment 1 ATPase and thin filament cooperativity. The role of tropomyosin flexibility and end-to-end interactions. *Biochemistry* **36**, 13449–13454 (1997).
39. Lorenz, M., Poole, K. J., Popp, D., Rosenbaum, G. & Holmes, K. C. An atomic model of the

- unregulated thin filament obtained by X-ray fiber diffraction on oriented actin-tropomyosin gels. *J. Mol. Biol.* **246**, 108–119 (1995).
40. McKillop, D. F. & Geeves, M. A. Regulation of the interaction between actin and myosin subfragment 1: evidence for three states of the thin filament. *Biophys. J.* **65**, 693–701 (1993).
 41. Fujita-Becker, S., Kluwe, L., Miegel, A., Maeda, K. & Maeda, Y. Reconstitution of rabbit skeletal muscle troponin from the recombinant subunits all expressed in and purified from *E. coli*. *J. Biochem. (Tokyo)* **114**, 438–444 (1993).
 42. Otwinoski, Z. M. W. Processing of X-ray diffraction data collected in oscillation mode (eds Carter, C. W. & Sweet, R. M.) *Methods Enzymol.* **276**, 307–326 (1997).
 43. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
 44. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. (eds Carter, C. W. & Sweet, R. M.) *Methods Enzymol.* **276**, 472–494 (1997).
 45. Abrahams, J. P. & Leslie, A. G. W. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D* **52**, 30–42 (1996).
 46. Roussel, A. & Cambillau, C. *TURBO-FRODO Manual* (AFMB-CNRS, Marseille, 1996).
 47. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
 48. CCP4 The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
 49. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *Acta Crystallogr. D* **24**, 946–950 (1991).
 50. Merritt, E. A. & Murphy, E. M. P. *Raster3D Version 2.0—a program for photorealistic molecular graphics* *Acta Crystallogr. D* **50**, 869–873 (1994).
- Supplementary Information** accompanies the paper on www.nature.com/nature.
- Acknowledgements** We thank Y. Kawano, S. Adachi, S.-Y. Park, M. Kawamoto and K. Miura for technical help at the beam lines of SPring-8. We also thank S. Ebashi, I. Ohtsuki, F. Oosawa, K. Maruyama and T. Nitta for continuous support and encouragement throughout this work. This work was supported in part by Matsushita Electric Industrial, and by the Special Coordination Funds from the Ministry of Education, Culture, Sports, Science and Technology, Japan. We dedicate this paper to S. Ebashi.
- Competing interests statement** The authors declare that they have no competing financial interests.
- Correspondence** and requests for materials should be addressed to S.T. (stakeda@ri.ncvc.go.jp) or Y.M. (ymaeda@spring8.or.jp). Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 1J1D for Tn46K and 1J1E for Tn52K.

Nuclear-powered millisecond pulsars and the maximum spin frequency of neutron stars

Deepto Chakrabarty^{*,†}, Edward H. Morgan^{*}, Michael P. Muno^{*}, Duncan K. Galloway^{*}, Rudy Wijngaards[‡], Michiel van der Klis[§] & Craig B. Markwardt^{||¶}

^{*} Department of Physics and Center for Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[†] Kavli Institute for Theoretical Physics, University of California, Santa Barbara, California 93106, USA

[‡] School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS, UK

[§] Astronomical Institute "Anton Pannekoek" and Center for High-Energy Astrophysics, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands

^{||} Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA

[¶] Laboratory for High Energy Astrophysics, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

Millisecond pulsars are neutron stars that are thought to have been spun-up by mass accretion from a stellar companion¹. It is not known whether there is a natural brake for this process, or if it continues until the centrifugal breakup limit is reached at submillisecond periods. Many neutron stars that are accreting mass from a companion star exhibit thermonuclear X-ray bursts that last tens of seconds, caused by unstable nuclear burning on their surfaces². Millisecond-period brightness oscillations during bursts from ten neutron stars (as distinct from other rapid X-ray variability that is also observed^{3,4}) are thought to measure the stellar spin^{2,3}, but direct proof of a rotational origin has been lacking. Here we report the detection of burst oscillations at the known spin frequency of an accreting millisecond pulsar, and we show that these oscillations always have the same rotational phase. This firmly establishes burst oscillations as nuclear-powered pulsations tracing the spin of accreting neutron stars, corroborating earlier evidence^{5,6}. The distribution of spin frequencies of the 11 nuclear-powered pulsars cuts off well below the breakup frequency for most neutron-star models, supporting theoretical predictions that gravitational radiation losses can limit accretion torques in spinning up millisecond pulsars^{7–9}.

The millisecond oscillations observed during X-ray bursts are not perfectly coherent, but usually drift in frequency by several hertz over the course of a burst, generally reaching an asymptotic maximum frequency that is repeatable in a given neutron star². This frequency drift has been interpreted as arising from angular momentum conservation in a decoupled surface burning layer that expands and contracts during the burst, so that the asymptotic frequency is the stellar spin frequency^{10,11}. A puzzle in this picture is why the oscillation persists late in the burst, well after the nuclear burning has spread over the entire star. Also, in most of these neutron stars, unexplained pairs of kilohertz quasi-periodic oscillations (kHz QPOs) are observed in the non-burst X-ray emission, with the QPO separation frequency approximately equal to either the burst oscillation frequency or half this value³, introducing a further puzzle (see ref. 4 for new insight into this problem).

We have observed the transient X-ray source SAX J1808.4–3658, which has been detected in four outbursts since its discovery (September 1996, April 1998, January 2000 and October 2002), each outburst lasting several weeks. It was previously established that this source is a weakly magnetized ($<10^{10}$ G), rapidly rotating (401 Hz), accreting neutron star in a 2-h low-mass X-ray binary^{12–14}.

We observed it during its 2002 X-ray outburst for about 700,000 s between 15 October and 26 November using the Proportional Counter Array (PCA) on the Rossi X-Ray Timing Explorer (RXTE). As in the two previous outbursts observed by RXTE, persistent 401-Hz accretion-powered X-ray pulsations¹³ modulated by the 2-h binary orbit¹⁴ were detected throughout our observations, with a fractional r.m.s. amplitude of 3–5%. The pulsar spin frequency derived from these data was approximately 400.97521 Hz at the start of the outburst, with a mean spin-down rate of about 2×10^{-13} Hz s⁻¹; a detailed discussion of the pulsar's spin evolution will be presented elsewhere. Four thermonuclear X-ray bursts were also detected on 15, 17, 18 and 19 October; these are among the brightest X-ray bursts ever observed by RXTE from any neutron star. Previous analysis of observations from the 1996 outburst of SAX J1808.4–3658 with the BeppoSAX satellite yielded

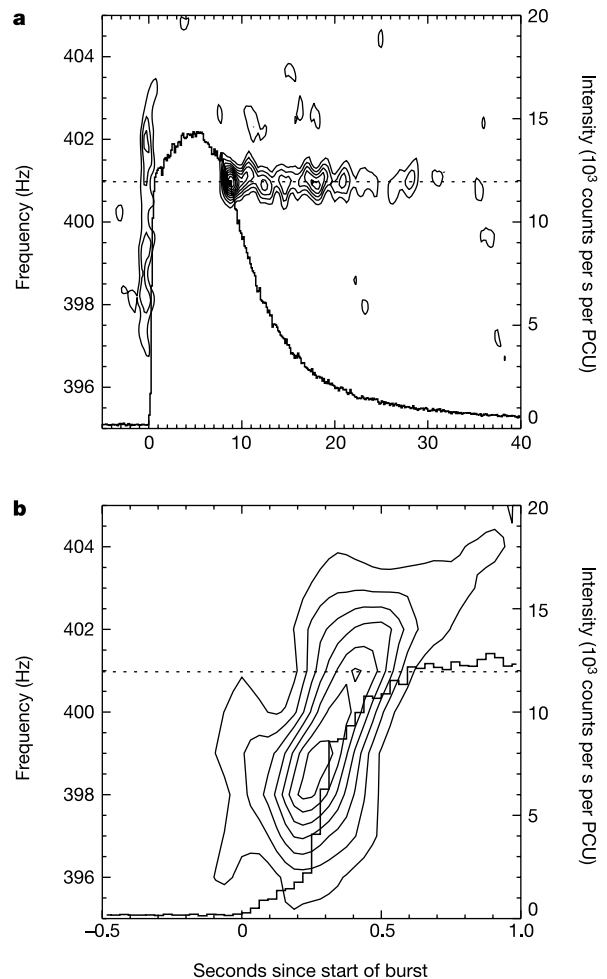


Figure 1 Dynamic power spectra of millisecond oscillations in an X-ray burst on 18 October 2002 from SAX J1808.4–3658. **a**, We analysed the 2–60-keV PCA data binned at 122- μ s resolution with no energy resolution, with the bin times corrected to the binary centre-of-mass frame. We computed overlapping, oversampled Fourier power spectra of 2-s duration spaced at 0.25-s intervals. The contours show Fourier power levels as a function of frequency and time, and the solid histogram shows the X-ray count rate per proportional counter unit (PCU). The horizontal dotted line shows the (known) pulsar spin frequency. A rapidly drifting oscillation is detected during the burst rise, and a stationary oscillation at nearly the spin frequency is detected in the burst tail. The n th contour level has a single-trial probability of 0.02ⁿ of occurring owing to noise fluctuations. **b**, A close-up view of the burst rise, computed using power spectra of 0.25-s duration at 0.03125-s intervals. The smooth frequency drift is more clearly apparent here.

a marginal detection of a 400 ± 2 Hz oscillation during a bright X-ray burst¹⁵.

Strong millisecond oscillations around 401 Hz were clearly detected in all four bursts in the October 2002 data, with a fractional r.m.s. amplitude of 3–5% and very similar characteristics in each burst (Fig. 1). First, a rapidly drifting oscillation was detected during the burst rise in the 397–403 Hz range. Second, no oscillations were detected ($<1.5\%$ amplitude) during the peak luminosity phase of the burst, when the photosphere is driven away from the star by radiation pressure; this behaviour is also typical of other burst oscillation sources¹⁶. Finally, a strong oscillation reappeared during the cooling phase of the burst at a constant frequency nearly equal to the pulsar spin frequency, but exceeding it by an average of 6 ± 1 mHz in the three fully sampled bursts. The oscillation amplitudes are comparable to those of the persistent (non-burst) accretion-powered pulsations, even though the burst emission is considerably brighter than the persistent emission, indicating that the burst flux itself must be pulsed. The burst oscillations are sinusoidal in shape, with a 2σ upper limit of $<0.5\%$ on the fractional r.m.s. amplitude of an 802-Hz harmonic, compared to a measured mean harmonic amplitude of 0.4% for the persistent pulsations.

The observed frequency drift demonstrates that this is a similar phenomenon to the burst oscillations observed in ten other neutron stars, although the oscillations in this neutron star have some unusual traits. The frequency drift is among the largest observed in any neutron star^{17,18} ($\Delta\nu \approx 5$ Hz, $\Delta\nu/\nu \approx 1\%$). Also, the drift timescale is an order of magnitude faster than in the other neutron stars¹⁶, and the maximum oscillation frequency is reached during the burst rise, inconsistent with angular momentum conservation in a cooling, contracting shell. In fact, the oscillation overshoots the spin frequency during the burst rise. The rapid frequency drift may be an indication that SAX J1808.4–3658 has a stronger magnetic

field than the other neutron stars, as a sufficiently strong field will suppress rotational shearing in the burning layer¹¹ and may act as a restoring force.

This magnetic field argument is particularly appealing, given that SAX J1808.4–3658 is the only burst oscillation source (and one of only four neutron stars in low-mass X-ray binaries out of over 70) that shows persistent pulsations in its non-burst emission^{19–21}. The absence of persistent pulsations in most of these neutron stars suggests that they lack a sufficiently strong field for the accretion flow to be magnetically channelled. Indeed, it has been proposed that the absence of persistent millisecond pulsations from most neutron stars in low-mass X-ray binaries is due to diamagnetic screening of the neutron-star magnetic field by freshly accreted material for the typical range of mass accretion rates²² ($\dot{M} \geq 10^{-10} M_{\odot} \text{ yr}^{-1}$, where $M_{\odot}$ is the solar mass). In this context, we note that all four persistent millisecond X-ray pulsars lie at the low end ($\dot{M} \approx 10^{-11} M_{\odot} \text{ yr}^{-1}$) of the mean $\dot{M}$ distribution for low-mass X-ray binaries. If this explanation is correct, then most burst oscillation sources are unlikely to show persistent pulsations.

We can measure the rotational phase of the oscillations in the burst tails. For each burst, we determined a rotational ephemeris using the persistent (accretion-powered) pulsations in the data for a few thousand seconds before the burst, and compared this to the phase (epoch of maximum intensity) of the oscillation in the burst tail. The burst tail oscillations all had the same rotational phase (within $\pm 6\%$), and were roughly phase-aligned with the non-burst pulsations, leading them by an average of 11% (Fig. 2). These measurements will help resolve the puzzle of why oscillations persist in the burst tail, when the nuclear burning has presumably enveloped the entire star. The phase locking indicates that the oscillations are associated with the stellar surface, and suggests that the emitting ‘hotspot’ has a nearly fixed orientation with respect to the pulsar’s magnetic axis. The 11% offset is consistent with the phase drift accumulated over the duration of the burst tails owing to the slight ($<0.002\%$) frequency difference of the burst oscillations and the non-burst pulsations, but this requires an initial phase alignment in the burst tail followed by slight motion of the hotspot²³.

Our observations clearly show that brightness oscillations in the tails of thermonuclear X-ray bursts are a nearly exact tracer of stellar spin, establishing these neutron stars as nuclear-powered pulsars. (Other recent observations show that the kHz QPO frequency separation in this neutron star is half the spin frequency⁴, verifying that the burst oscillation traces the spin and not a harmonic.) This provides a powerful tool for studying the most rapidly rotating neutron stars. It is believed that accretion from a binary companion is required in order to spin-up old neutron stars to millisecond periods¹. For sustained accretion in the absence of another angular momentum sink, this process is limited only by the neutron-star breakup frequency—which is up to 3 kHz, depending upon the unknown nuclear equation of state^{24,25}. It has been difficult to determine the actual spin distribution of millisecond pulsars, as there have been strong observational selection effects against detecting very short period millisecond radio pulsars, and persistent millisecond X-ray pulsars are rare. However, RXTE has no significant selection effects against detecting burst oscillations at frequencies well above 1 kHz, so the nuclear-powered pulsars are ideal for probing the fast end of the pulsar spin distribution.

The spin frequencies of the 11 nuclear-powered pulsars lie between 270 Hz and 619 Hz, and within that range are marginally consistent with a uniform distribution. However, the absence of spins above 619 Hz is significant. If we make the simple assumption of a uniform spin distribution in the range $[\nu_{\text{low}}, \nu_{\text{high}}]$, and consider a sample of N pulsars with spin frequencies ν_j , then a bayesian analysis yields the probability density for ν_{high} (within a normalization factor),

$$p(\nu_{\text{high}}) = [\nu_{\text{high}} - \min(\nu_j)]^{1-N} - \nu_{\text{high}}^{1-N} \quad (1)$$

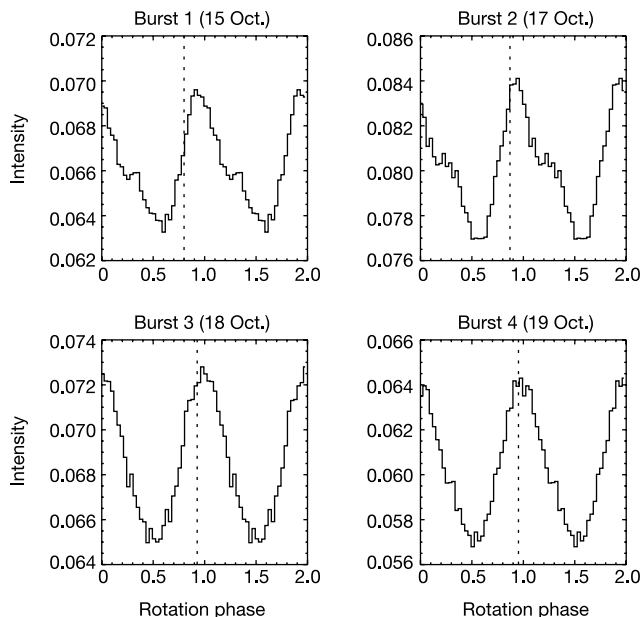


Figure 2 Relative phase of the burst tail oscillations in SAX J1808.4–3658, and the persistent accretion-powered pulsations just before the burst, for all four bursts. The histograms show the profile of the persistent pulsations, and the dashed lines indicate the rotational phase of the burst tail oscillations. The oscillations all have nearly the same rotational phase, and slightly lead the persistent pulsations.

for $\nu_{\text{high}} \geq \max(\nu_j)$, where we have made an *a priori* allowance for a $0 < \nu < 3$ kHz range. For the 11 observed spins, this yields an upper limit of $\nu_{\text{high}} < 760$ Hz (95% confidence) on the maximum spin frequency, which is also consistent with the fastest known (641 Hz) millisecond radio pulsar²⁶. This is well below the breakup frequency for nearly all models of rapidly rotating neutron stars (except for those that have an extremely stiff equation of state and contain a pion/kaon condensate in the core, and then only for $M < 1.5 M_{\odot}$)²⁷. Some mechanism clearly acts to halt pulsar spin-up while accretion is still active.

Steady magnetic accretion torques should drive an accreting pulsar to an equilibrium spin frequency that depends upon $\dot{M}$ and the surface dipole magnetic field B . Taking the $\dot{M}$ range observed in the nuclear-powered pulsars and the disk-magnetosphere interaction models relevant for weakly magnetic neutron stars²⁸, magnetic spin equilibrium can account for the observed spin distribution if all these systems have $B \approx 10^8$ G, consistent with the field inferred in SAX J1808.4–3658 (ref. 28) and in the millisecond radio pulsars²⁹. However, fields this strong should be dynamically important for the accretion flow, making it difficult to understand the lack of persistent pulsations in the non-burst emission of all the nuclear-powered pulsars besides SAX J1808.4–3658, and instead suggesting a broader range of field strengths. Alternatively, several authors have shown that gravitational radiation can carry away substantial angular momentum from accreting neutron stars, driven either by the excitation of an r-mode instability in the neutron-star core^{7,9} or by a rotating, accretion-induced crustal quadrupole moment⁸. These losses can balance accretion torques for the relevant ranges of ν and $\dot{M}$, and the predicted gravitational radiation strengths are near the detection threshold for planned gravitational wave interferometers, especially in the case where an X-ray timing ephemeris is available³⁰. □

Received 19 March; accepted 21 May 2003; doi:10.1038/nature01732.

- Alpar, M. A., Cheng, A. F., Ruderman, M. A. & Shaham, J. A new class of radio pulsars. *Nature* **300**, 728–730 (1982).
- Strohmayer, T. E. & Bildsten, L. in *Compact Stellar X-Ray Sources* (eds Lewin, W. H. G. & van der Klis, M.) (Cambridge Univ. Press, in the press); preprint at (<http://arXiv.org/astro-ph/0301544>) 2003.
- van der Klis, M. Millisecond oscillations in X-ray binaries. *Annu. Rev. Astron. Astrophys.* **38**, 717–760 (2000).
- Wijnands, R. *et al.* Kilohertz quasi-periodic X-ray brightness fluctuations from an accreting millisecond pulsar. *Nature* (in the press).
- Strohmayer, T. E. *et al.* Millisecond variability from an accreting neutron star system. *Astrophys. J.* **469**, L9–L12 (1996).
- Strohmayer, T. E. & Markwardt, C. B. Evidence for a millisecond pulsar in 4U 1636–53 during a superburst. *Astrophys. J.* **577**, 337–345 (2002).
- Wagoner, R. V. Gravitational radiation from accreting neutron stars. *Astrophys. J.* **278**, 345–348 (1984).
- Bildsten, L. Gravitational radiation and rotating neutron stars. *Astrophys. J.* **501**, L89–L93 (1998).
- Andersson, N., Kokkotas, K. D. & Stergioulas, N. On the relevance of the r-mode instability for accreting neutron stars and white dwarfs. *Astrophys. J.* **516**, 307–314 (1999).
- Strohmayer, T. E., Jahoda, K., Giles, A. B. & Lee, U. Millisecond pulsations from a low-mass X-ray binary system in the Galactic center region. *Astrophys. J.* **486**, 355–362 (1997).
- Cumming, A. & Bildsten, L. Rotational evolution during type I X-ray bursts. *Astrophys. J.* **544**, 453–474 (2000).
- in't Zand, J. J. *et al.* Discovery of the X-ray transient SAX J1808.4–3658, a likely low-mass X-ray binary. *Astron. Astrophys.* **331**, L25–L28 (1998).
- Wijnands, R. & van der Klis, M. A millisecond pulsar in an X-ray binary system. *Nature* **394**, 344–346 (1998).
- Chakrabarty, D. & Morgan, E. H. The two-hour orbit of a binary millisecond X-ray pulsar. *Nature* **394**, 346–348 (1998).
- in't Zand, J. J. *et al.* The first outburst of SAX J1808.4–3658 revisited. *Astron. Astrophys.* **372**, 916–921 (2001).
- Muno, M. P., Chakrabarty, D., Galloway, D. K. & Psaltis, D. The frequency stability of millisecond oscillations in thermonuclear X-ray bursts. *Astrophys. J.* **580**, 1048–1059 (2002).
- Wijnands, R., Strohmayer, T. & Franco, L. M. Discovery of nearly coherent oscillations with a frequency of 567 Hz during type I X-ray bursts of the X-ray transient and eclipsing binary X1658–298. *Astrophys. J.* **549**, L71–L75 (2001).
- Galloway, D. K., Chakrabarty, D., Muno, M. P. & Savov, P. Discovery of a 270 hertz X-ray burst oscillation in the X-ray dipper 4U 1916–053. *Astrophys. J.* **549**, L85–L88 (2001).
- Markwardt, C. B., Swank, J. H., Strohmayer, T. E., in't Zand, J. J. M. & Marshall, F. E. Discovery of a second millisecond accreting pulsar: XTE J1751–305. *Astrophys. J.* **573**, L21–L24 (2002).
- Galloway, D. K., Chakrabarty, D., Morgan, E. H. & Remillard, R. A. Discovery of a high-latitude accreting millisecond pulsar in an ultracompact binary. *Astrophys. J.* **576**, L137–L140 (2002).
- Markwardt, C. B., Smith, E. & Swank, J. H. XTE J1807–294. *IAU Circ. No.* 8080 (2003).

- Cumming, A., Zweibel, E. & Bildsten, L. Magnetic screening in accreting neutron stars. *Astrophys. J.* **557**, 958–966 (2001).
- Spitkovsky, A., Levin, Y. & Ushomirsky, G. Propagation of thermonuclear flames on rapidly rotating neutron stars: Extreme weather during type I X-ray bursts. *Astrophys. J.* **566**, 1018–1038 (2002).
- Cook, G. B., Shapiro, S. L. & Teukolsky, S. A. Recycling pulsars to millisecond periods in general relativity. *Astrophys. J.* **421**, L117–L120 (1994).
- Haensel, P., Lasota, J. P. & Zdzunik, J. L. On the minimum period of uniformly rotating neutron stars. *Astron. Astrophys.* **344**, 151–153 (1999).
- Backer, D. C., Kulkarni, S. R., Heiles, C. E., Davis, M. M. & Goss, W. M. A millisecond pulsar. *Nature* **300**, 615–618 (1982).
- Cook, G. B., Shapiro, S. L. & Teukolsky, S. A. Rapidly rotating neutron stars in general relativity: Realistic equations of state. *Astrophys. J.* **424**, 823–845 (1994).
- Psaltis, D. & Chakrabarty, D. The disk-magnetosphere interaction in the accretion-powered millisecond pulsar SAX J1808.4–3658. *Astrophys. J.* **521**, 332–340 (1999).
- Camilo, F., Thorsett, S. E. & Kulkarni, S. R. The magnetic fields, ages, and original spin periods of millisecond pulsars. *Astrophys. J.* **421**, L15–L18 (1994).
- Bildsten, L. in *Radio Pulsars* (eds Bailes, M., Nice, D. J. & Thorsett, S. E.) (Astronomical Society of the Pacific, San Francisco, in the press); preprint at (<http://arXiv.org/astro-ph/0212004>) (2002).

Acknowledgements We thank L. Bildsten, F. Lamb, D. Psaltis and S. Thorsett for discussions. This work was supported in part by NASA and the Alfred P. Sloan Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.C. (deepto@space.mit.edu).

Quasi-periodic X-ray brightness fluctuations in an accreting millisecond pulsar

R. Wijnands*, M. van der Klis†, J. Homan‡, D. Chakrabarty§, C.B. Markwardt|| & E.H. Morgan§

* School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS, UK

† Astronomical Institute 'Anton Pannekoek', University of Amsterdam, and Centre for High-Energy Astrophysics, Kruislaan 403, 1098 SJ, Amsterdam, The Netherlands

‡ INAF-Osservatorio Astronomico di Brera, Via E. Bianchi 46, 23807 Merate LC, Italy

§ Center for Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

|| Laboratory for High Energy Astrophysics, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

The relativistic plasma flows onto neutron stars that are accreting material from stellar companions can be used to probe strong-field gravity as well as the physical conditions in the supranuclear-density interiors of neutron stars. Plasma inhomogeneities orbiting a few kilometres above the stars are observable as X-ray brightness fluctuations on the millisecond dynamical timescale of the flows^{1–3}. Two frequencies in the kilohertz range dominate these fluctuations: the twin kilohertz quasi-periodic oscillations (kHz QPOs). Competing models for the origins of these oscillations (based on orbital motions) all predict that they should be related to the stellar spin frequency^{4–10}, but tests have been difficult because the spins were not unambiguously known. Here we report the detection of kHz QPOs from a pulsar whose spin frequency is known. Our measurements establish a clear link between kHz QPOs and stellar spin, but one not predicted by any current model. A new approach to understanding kHz QPOs is now required. We suggest that a resonance between the spin and general relativistic orbital and epicyclic frequencies could provide the observed relation between QPOs and spin.

Twin kHz QPOs with frequencies between 300 and 1,300 Hz

occur in more than 20 accreting neutron stars³. In seven of these, during thermonuclear X-ray bursts caused by the explosion of material accumulated on the neutron-star surface, short-lived oscillations are observed with frequencies ν_{burst} in the range 270–620 Hz, approximately constant but different in each source, that are thought to be caused by the neutron-star spin¹¹. The frequency difference $\Delta\nu$ between the two kHz QPOs is close to either ν_{burst} or $\nu_{\text{burst}}/2$, depending on the source. This observed commensurability has given rise to beat-frequency models^{6,12–14}. Such models identify the higher-frequency (‘upper’) kHz QPO with the frequency of

orbital motion, closely around the neutron star, of the plasma at the inner edge of the accretion disk (within this radius no stable orbits exist and the matter plunges in), and the other (‘lower’) kHz QPO with a rotational beat interaction between the upper kHz QPO and the neutron-star spin. These models predict $\Delta\nu$ to be equal to the spin frequency ν_{spin} . In this interpretation, we have either $\nu_{\text{burst}} = \nu_{\text{spin}}$ or $\nu_{\text{burst}} = 2\nu_{\text{spin}}$, depending on whether one or two hotspots form on the star’s surface during the bursts. Testing this requires searching for harmonic structure in the burst oscillations, but searches¹⁵ are limited in terms of statistics as compared to the case of a true pulsar.

A problem for beat-frequency models is the fact that $\Delta\nu$ is neither constant^{16,17}, nor exactly equal^{18,19} to ν_{spin} (but possible solutions in the beat-frequency models do exist²⁰). This has motivated the proposal of several alternative models to explain the kHz QPOs^{7–10}, all of which, like beat-frequency models, use orbital motion as one of the observed frequencies, but which use another mechanism—such as general relativistic epicyclic motion, which is very fast in these extreme gravitational fields—to generate the other frequency. None of these models predict commensurability between spin and kHz QPO frequencies such as beat-frequency models do. It has been clear for some time that independent measurements of the neutron-star spin frequency could provide the definitive test of the beat-frequency idea. The discovery²¹ of the first accreting millisecond pulsar in 1998 raised the prospect of such a test, but until now no accreting millisecond pulsar has shown the anticipated kHz QPOs.

On 13 October 2002, a bright new outburst of the recurrent X-ray transient and accreting millisecond pulsar, SAX J1808.4–3658, started²². Using the Rossi X-ray Timing Explorer satellite, we obtained ~700 ks of data covering the outburst with daily observations from 15 October to 26 November 2002. We performed a series of fast Fourier transforms to search for kHz QPOs, and discovered several. One kHz QPO was monitored for 12 days (Fig. 1). Its frequency increased from 570 to 725 Hz between 15 and 16 October, then gradually decreased to a minimum of 280 Hz on 21 October, suddenly increased again to 400 Hz the next day, and then resumed its decrease, to 320 Hz on 26 October. After this the QPO was not detected, presumably owing to the limited statistics resulting from the drop in count rate (by a factor of about 4) by this stage of the outburst. On 16 October, a second kHz QPO was detected simultaneously at a frequency of $\Delta\nu = 195 \pm 6$ Hz below the first one (Fig. 1, top panel). In the remaining data, this second QPO was not detected with a significance greater than 2.5σ . Additional broad noise structures of marginal significance are sometimes present around 350, 220 and 120 Hz.

The frequencies, strengths and coherences of these two kHz QPOs in SAX J1808.4–3658, as well as the variations in their parameters as a function of source luminosity, are very similar to those of the twin kHz QPOs observed in those neutron-star low-mass binaries for which no pulsations are seen in their persistent emission (the non-pulsing sources³), so we interpret this as the same phenomenon. The ~195-Hz frequency difference between the two QPOs is the lowest known (in other systems $\Delta\nu$ ranges between 225 and 350 Hz)³. It is far below the 401-Hz spin frequency predicted in simple beat-frequency models^{6,12}, yet there is a clear commensurability: $\Delta\nu$ is consistent with $\nu_{\text{spin}}/2$. This is reminiscent of those non-pulsing sources where $\Delta\nu$ is near $\nu_{\text{burst}}/2$. Hence our results support the spin interpretation of the burst oscillations in those sources, as well as the suspicion that ν_{burst} is always near ν_{spin} (and never twice that). Moreover, as reported in a companion paper²³, in our observations we also detected four thermonuclear bursts in which we discovered burst oscillations at the neutron-star spin frequency. This discovery further strengthens the idea that the burst oscillations are always at the stellar spin frequency.

We now demonstrate that, unless we are wrong about the spin frequency of this pulsar, which seems highly unlikely, our obser-

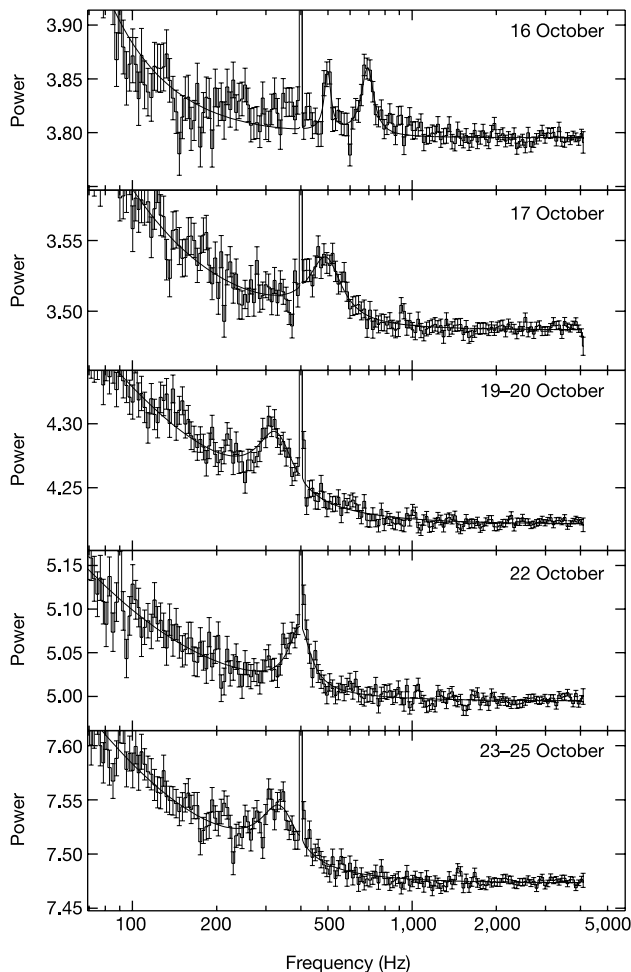


Figure 1 The power-density spectra obtained for SAX J1808.4–3658 during its 2002 outburst. The power is given in units of $(\text{r.m.s./mean})^2/(\text{Hz} \times 10^{-3})$. These spectra show the presence of the pulsations at 401 Hz, the QPOs, and broadband noise below a few hundred hertz. The date the data were obtained is given in each panel. All data during those days (for the energy range 3–60 keV) were used to create the power-density spectra, except for that for 16 October for which only the data between 0:53 and 10:00 UTC were used and for the energy range 3.3–39 keV. The properties of the QPO were obtained by fitting the power-density spectra with a constant for the Poisson level (clearly visible at frequencies $>1,000$ Hz), and three lorentzian functions to fit the continuum at low frequencies, the pulsation frequency at 401 Hz, and the upper kHz QPO. On 16 October, two kHz QPOs are visible with frequencies of 499 ± 4 Hz and 694 ± 4 Hz, fractional r.m.s. amplitudes (3.3–39 keV) of $5.2 \pm 0.5\%$ (5σ) and $9.1 \pm 0.5\%$ (9σ) and widths (FWHM) of 35 ± 12 Hz and 82 ± 15 Hz, which we identify as the lower and the upper kHz QPO, respectively. In the remaining observations one kHz QPO is detected which we identify as the upper kHz QPO based on its fractional r.m.s. amplitude of 8–10%, FWHM of about 100 Hz, and smooth relation of its frequency to that of QPOs and noise features below 100 Hz. No systematic differences in its other properties were found when this QPO varied in frequency between 725 Hz and 320 Hz.

vations falsify current beat-frequency models, yet so pose a severe challenge to all other current kHz QPO models. In beat-frequency models^{5,6,24}, the spin-orbit beat interaction occurs at a frequency $\nu_{\text{beat}} = n(\nu_{\text{orbit}} - \nu_{\text{spin}})$, where ν_{orbit} is the orbital frequency, and the positive integer n is a symmetry factor accounting for multiple symmetrically located spin-orbit interaction sites (for example, $n = 2$ for two magnetic poles). Simple beat-frequency models have $n = 1$ and the two kHz QPOs at ν_{beat} and ν_{orbit} , so that $\Delta\nu = \nu_{\text{orbit}} - \nu_{\text{beat}} = \nu_{\text{spin}}$. As noted, this is clearly inconsistent with our observations. In general, n is the least common multiple (LCM) of the number of stellar and the number of orbital interaction sites: $n = \text{LCM}(n_{\text{spin}}, n_{\text{orbit}})$. Whatever the value of these numbers, and even allowing the observed orbital QPO to be at either ν_{orbit} or $n_{\text{orbit}}\nu_{\text{orbit}}$, there is no configuration predicting the observed $\Delta\nu = \nu_{\text{spin}}/2$.

The only solution consistent with current beat-frequency models is that ν_{spin} is half the observed 401-Hz pulsar frequency (that is, two hotspots) and $n = 1$ (that is, only one pulsar beam or magnetic pole interacts with the orbit). However, the very stringent amplitude upper limit of 0.014% root mean square (r.m.s.; E.H.M. *et al.*, manuscript in preparation) on any coherent signal at 200.5 Hz (only 0.38% of the signal at 401 Hz) makes this solution very unlikely: a fine-tuned special geometry would be required to hide the true spin frequency that well. We note that the sonic point beat-frequency model⁶ predicts a spectrum of many additional weaker QPO peaks (higher-order sidebands), one of which (for example, $\text{abs}(\nu_{\text{upper}} - 3\nu_{\text{spin}})$ or, a particularly close match, $\text{abs}(3\nu_{\text{lower}} - 2\nu_{\text{spin}})$, where ν_{upper} and ν_{lower} are the upper and lower kHz QPO frequency, respectively) could be picked out and identified with a QPO that we observe. However, this would not explain why, contrary to predictions, this peak is stronger—rather than much weaker—than the main peaks. In our view, the coincidence between $\Delta\nu$ and $\nu_{\text{burst}}/2$ in several other sources strongly suggests instead that the relation between kHz QPOs and neutron-star spin is one where (in SAX J1808.4–3658 and in those several other sources) the peaks are separated by $\nu_{\text{spin}}/2$. This has not been predicted by any model.

Inspired by the relativistic resonance models of ref. 9 (see also refs 25–27 for related suggestions), we note that there are particular radii in the disk where the difference between the general relativistic

orbital frequency (ν_{ϕ}) and radial frequency (ν_r ; note that $\nu_{\phi} - \nu_r$ is the periastron precession frequency) is equal to ν_{spin} and $\nu_{\text{spin}}/2$. For a 1.4-solar-mass Schwarzschild geometry, these radii are ~ 18 and ~ 23 km, respectively (the best match to the observed frequencies in fact occurs at 1.1 solar masses, but see below). If a radiatively or magnetically mediated spin-orbit interaction were to set up a resonance at one of those radii leading to kHz QPOs at ν_{ϕ} and ν_r , this could explain the observed frequency commensurabilities with no involvement of a beat frequency. Whether $\Delta\nu$ is ν_{spin} or $\nu_{\text{spin}}/2$ might depend on whether or not the relevant radius is within the inner edge of the disk; in support of this, we note that all four systems where $\Delta\nu \approx \nu_{\text{spin}}/2$ are those where ν_{spin} exceeds 400 Hz, and vice versa for the remaining three cases. That the frequencies can shift and the observed commensurabilities are often not exact might be related to a mechanism similar to that which makes the resonances occur off the naively expected integer ratios in the numerical calculations of ref. 28. We note that this combination of ideas (adding the spin as an ingredient in the relativistic resonance model) can lead to a single framework within which the apparently discrepant frequency commensurabilities observed in black holes (which take the form of integer ratios) as well as neutron stars can both be understood in terms of resonances at privileged radii in an accretion disk, namely those radii where the general relativistic epicyclic frequencies (and, in the case of the neutron-star systems, spin) are commensurate.

A surprising third QPO (Fig. 2) near 410 Hz was detected on four occasions, just above the 401-Hz pulse frequency. Its frequency increased from 409.3 ± 0.7 to 413.5 ± 0.2 Hz while the frequency of the upper kHz QPO increased from 328 ± 4 to 390 ± 5 Hz. This third kHz QPO is a phenomenon not previously seen, and is probably related to the pulsating nature of SAX J1808.4–3658. To some extent, it resembles the sideband to the lower kHz QPO observed²⁹ in three non-pulsing kHz QPO sources: the 410-Hz QPO might be a sideband to the pulsation with a similar underlying mechanism. These lower kHz QPO sidebands were suggested²⁹ to be due to Lense-Thirring precession, a predicted—but not yet observed—general relativistic wobble of the orbital plane (nodal precession). The precession frequency is predicted³⁰ to be quadratically related to the orbital frequency itself, and in the case of both the QPO sidebands and the ~ 410 -Hz QPO this is consistent with

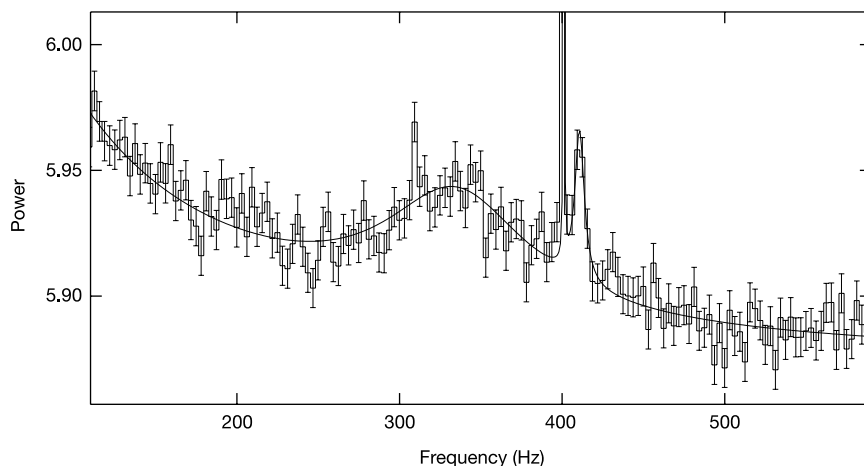


Figure 2 The power-density spectrum obtained for SAX J1808.4–3658 using the combined data from 18–19 October and 22–26 October. The power is given in units of $(\text{r.m.s.}/\text{mean})^2/(\text{Hz} \times 10^{-3})$. This spectrum is for the photon energy range 3–60 keV, and shows the presence of the pulsations at 401 Hz, the upper kHz QPO, the ~ 410 -Hz QPO, and broadband noise below a few hundred hertz. The spectrum was fitted with the same function as described in Fig. 1, but with an extra lorentzian function to fit

the extra QPO at ~ 410 Hz. The extra QPO could be detected significantly on four occasions: in the combined 18–19 October data, the 22 October data, the combined 23–24 October data, and the combined 25–26 October data. Combining all those data, the QPO was significant at a 6.5σ level. This QPO had an r.m.s amplitude of $\sim 2.5\%$ (for the energy range 3–60 keV) and its width (FWHM) was between 2 and 8 Hz.

observations. However, as Lense-Thirring precession is prograde, a beat between a pulsar beam and a precessing orbit would produce a sideband below, rather than the observed sideband above, the pulsar frequency. Another resonance, perhaps at the radius where the general relativistic vertical epicyclic frequency matches ν_{spin} , might be considered as an explanation of this phenomenon. □

Received 19 March; accepted 25 May 2003; doi:10.1038/nature01754.

1. Svartsman, V. F. Halos around black holes. *Sov. Astron.* **15**, 377–384 (1971).
2. Van der Klis, M. *et al.* Intensity dependent quasi-periodic oscillations in the X-ray flux of GX 5-1. *Nature* **316**, 225–230 (1985).
3. Van der Klis, M. Millisecond oscillations in X-ray binaries. *Annu. Rev. Astron. Astrophys.* **38**, 717–760 (2000).
4. Alpar, M. A. & Shaham, J. Is GX 5-1 a millisecond pulsar? *Nature* **316**, 239–241 (1985).
5. Lamb, F. K., Shibasaki, N., Alpar, M. A. & Shaham, J. Quasi-periodic oscillations in bright galactic-bulge X-ray sources. *Nature* **317**, 681–687 (1985).
6. Miller, C. M., Lamb, F. K. & Psaltis, D. Sonic-point model of kilohertz quasi-periodic brightness oscillations in low-mass X-ray binaries. *Astrophys. J.* **508**, 791–830 (1998).
7. Stella, L. & Vietri, M. KHz quasi-periodic oscillations in low-mass X-ray binaries as probes of general relativity in the strong-field regime. *Phys. Rev. Lett.* **82**, 17–20 (1999).
8. Titarchuk, L., Lapidus, I. & Muslimov, A. Mechanisms for high-frequency quasi-periodic oscillations in neutron stars and black hole binaries. *Astrophys. J.* **449**, 315–328 (1998).
9. Kluzniak, W. & Abramowicz, M. A. The physics of kHz QPOs—strong gravity's coupled anharmonic oscillators. *Astrophys. J. Lett.* (submitted); preprint at (<http://xxx.lanl.gov/astro-ph/0105057>) (2001).
10. Psaltis, D. & Norman, C. On the origin of quasi-periodic oscillations and broad-band noise in accreting neutron stars and black holes. *Astrophys. J.* (submitted); preprint at (<http://xxx.lanl.gov/astro-ph/0001391>) (2000).
11. Strohmayer, T. & Bildsten, L. New views of thermonuclear bursts. In *Compact Stellar X-ray Sources* (eds Lewin, W. H. G. & van der Klis, M.) (Cambridge Univ. Press, in the press); preprint at (<http://xxx.lanl.gov/astro-ph/0301544>) (2003).
12. Strohmayer, T. E. *et al.* Millisecond X-ray variability from an accreting neutron star system. *Astrophys. J.* **469**, L9–L12 (1996).
13. Cui, W. On the disappearance of kilohertz quasi-periodic oscillations at high mass accretion rate in low-mass X-ray binaries. *Astrophys. J.* **534**, L31–L34 (2000).
14. Campana, S. Kilohertz quasi-periodic oscillations in low-mass X-ray binary sources and their relation to the neutron star magnetic field. *Astrophys. J.* **534**, L79–L82 (2000).
15. Muno, M. P., Özel, F. & Chakrabarty, D. The amplitude evolution and harmonic content of millisecond oscillations in thermonuclear X-ray bursts. *Astrophys. J.* **581**, 550–561 (2002).
16. Van der Klis, M., Wijnands, R. A. D., Horne, K. & Chen, W. Kilohertz quasi-periodic oscillation peak separation is not constant in Scorpius X-1. *Astrophys. J.* **481**, L97–L101 (1997).
17. Méndez, M. *et al.* Kilohertz quasi-periodic oscillation peak separation is not constant in the atoll source 4U 1608-52. *Astrophys. J.* **505**, L23–L26 (1998).
18. Méndez, M., van der Klis, M. & van Paradijs, J. Difference frequency of kilohertz QPOs not equal to half the burst oscillation frequency in 4U 1636-53. *Astrophys. J.* **506**, L117–L119 (1998).
19. Jonker, P. G., Méndez, M. & van der Klis, M. Kilohertz quasi-periodic oscillations difference frequency exceeds inferred spin frequency in 4U 1636-53. *Mon. Not. R. Astron. Soc.* **336**, L1–L5 (2002).
20. Lamb, F. K. & Miller, M. C. Changing frequency separation of kilohertz quasi-periodic oscillations in the sonic-point beat-frequency model. *Astrophys. J.* **554**, 1210–1215 (2001).
21. Wijnands, R. & van der Klis, M. A millisecond pulsar in an X-ray binary system. *Nature* **394**, 344–346 (1998).
22. Markwardt, C. B., Miller, J. M. & Wijnands, R. SAX J1808.4-3658. *IAU Circ. No.* 7993 (2002).
23. Chakrabarty, D. *et al.* Nuclear-powered millisecond pulsars and the maximum spin frequency of neutron stars. *Nature* (this issue).
24. Shibasaki, N. & Lamb, F. K. Power spectra of quasi-periodic oscillations in luminous X-ray stars. *Astrophys. J.* **318**, 767–785 (1987).
25. Psaltis, D. Models of quasi-periodic variability in neutron stars and black holes. *Adv. Space Res.* **555**, 786–800 (2001).
26. Stella, L. In *X-ray Astronomy: Stellar Endpoints, AGN, and the Diffuse X-ray Background* (eds White, N. E., Malaguti, G. & Palumbo, G. G. C.) 365–375 (AIP Conference Proceedings Vol. 599, Melville, New York, 2001).
27. van der Klis, M. in *XEU—Studying the Evolution of the Hot Universe* (eds Hasinger, G., Boller, Th. & Parmar, A.) 354–362 (MPE Report 281, 2002).
28. Abramowicz, M. A., Karas, V., Kluzniak, W., Lee, W. H. & Rebusco, P. Non-linear resonance in nearly geodesic motion in low-mass X-ray binaries. *Publ. Astron. Soc. Jpn* **55**, 467–471 (2003).
29. Jonker, P. G., Méndez, M. & van der Klis, M. Discovery of a new, third kilohertz quasi-periodic oscillations in 4U 1608-52, 4U 1728-34, and 4U 1636-53: Sidebands to the lower kilohertz quasi-periodic oscillation? *Astrophys. J.* **540**, L29–L32 (2000).
30. Stella, L. & Vietri, M. Lense-Thirring precession and quasi-periodic oscillations in low-mass X-ray binaries. *Astrophys. J.* **492**, L59–L62 (1998).

Acknowledgements We thank J. Swank and E. Smith for their efforts in scheduling our RXTE observations of SAX J1808.4–3658, and D. Psaltis for discussions. M.v.d.K. was supported by NWO.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.W. (radw@st-andrews.ac.uk.).

Creation of ultracold molecules from a Fermi gas of atoms

Cindy A. Regal*, Christopher Ticknor*, John L. Bohn* & Deborah S. Jin†

* JILA, National Institute of Standards and Technology and Department of Physics, University of Colorado, Boulder, Colorado 80309-0440, USA

† Quantum Physics Division, National Institute of Standards and Technology, Boulder, Colorado 80309-0440, USA

Following the realization of Bose–Einstein condensates in atomic gases, an experimental challenge is the production of molecular gases in the quantum regime. A promising approach is to create the molecular gas directly from an ultracold atomic gas; for example, bosonic atoms in a Bose–Einstein condensate have been coupled to electronic ground-state molecules through photo-association¹ or a magnetic field Feshbach resonance². The availability of atomic Fermi gases offers the prospect of coupling fermionic atoms to bosonic molecules, thus altering the quantum statistics of the system. Such a coupling would be closely related to the pairing mechanism in a fermionic superfluid, predicted to occur near a Feshbach resonance^{3,4}. Here we report the creation and quantitative characterization of ultracold ⁴⁰K₂ molecules. Starting with a quantum degenerate Fermi gas of atoms at a temperature of less than 150 nK, we scan the system over a Feshbach resonance to create adiabatically more than 250,000 trapped molecules; these can be converted back to atoms by reversing the scan. The small binding energy of the molecules is controlled by detuning the magnetic field away from the Feshbach resonance, and can be varied over a wide range. We directly detect these weakly bound molecules through their radio-frequency photodissociation spectra; these probe the molecular wavefunction, and yield binding energies that are consistent with theory.

Feshbach resonances occur when the collision energy of two free atoms coincides with that of a quasi-bound molecular state^{5–7}. By varying the strength of an external magnetic field, experimenters can tune the relative atom–molecule energy through the Zeeman effect. This enables control over the strength of cold atom interactions, characterized by the *s*-wave scattering length *a*, as well as whether these interactions are effectively repulsive (*a* > 0) or attractive (*a* < 0). These resonances have been used to tune the interaction strength between atoms for both Bose and Fermi gases^{8–16}.

Another possible use of a Feshbach resonance is to controllably convert atoms into molecules. The energy of the molecular state associated with the Feshbach resonance coincides with the free atom threshold at the resonance peak, and the binding energy of the molecule varies smoothly with magnetic field on the repulsive side of the resonance (*a* > 0). Thus, one would expect that atoms could be coupled to molecules by ramping the detuning from the resonance using a time-dependent magnetic field^{17–19}. For example, ramping the magnetic field through a Feshbach resonance resulted in large losses in a ²³Na Bose–Einstein condensate (BEC)²⁰. Further, coherent oscillations between atoms and molecules in a BEC were observed following magnetic field pulses on the repulsive side of a ⁸⁵Rb Feshbach resonance². Here we report the efficient creation of bosonic molecules from fermionic ⁴⁰K atoms using a magnetic field ramp across a Feshbach resonance. We present clear evidence of diatomic molecule formation through direct, spectroscopic detection of these molecules.

The experiments reported here use previously developed techniques for cooling and spin state manipulation of ⁴⁰K (refs 10, 15, 21). Because of the quantum statistics of fermions, a mixture of two components—for example, atoms in different internal spin states—

is required to have *s*-wave interactions in the ultracold gas. With a total atomic spin $f = 9/2$ in its lowest hyperfine ground state, ^{40}K has ten available Zeeman spin-states m_f , where m_f is the spin projection quantum number.

Mixtures of atoms in two of these spin states are used in evaporative cooling of the gas, first in a magnetic trap and then in a far-off resonance optical dipole trap. The optical trapping potential has the distinct ability to trap atoms in any spin state as well as any molecules created from these atoms. For these experiments, the optical trap is characterized by radial frequencies ranging from $\nu_r = 215$ to 276 Hz , with the trap aspect ratio, ν_r/ν_z , fixed at 70 ± 20 . The temperature T of these two-component gases, measured in units of the Fermi temperature T_F , ranges from $T/T_F = 0.13$ to 0.33 . This degree of quantum degeneracy is near the lowest ever demonstrated in a Fermi gas of atoms^{14,15}.

Experiments are initiated by preparing atoms in a nearly equal, incoherent mixture of the $m_f = -5/2$ and $m_f = -9/2$ spin states. With these states, we access a previously reported Feshbach resonance located at a magnetic field of $224.21 \pm 0.05\text{ G}$ (Fig. 1b)¹⁵. A time-dependent current through an auxiliary coil provides magnetic field ramps near the resonance. Starting from a magnetic field of 227.81 G , the field is ramped at a rate of $40\text{ }\mu\text{s G}^{-1}$ across the resonance to various final values. The number of atoms remaining following the ramp is determined from an absorption image of the cloud (at $\sim 4\text{ G}$) after expansion from the optical trap. As the light used for these images is resonant with the atomic transition, but not with any molecular transitions, we selectively detect only the atoms. In Fig. 1a we present the observed atom number as a function of the

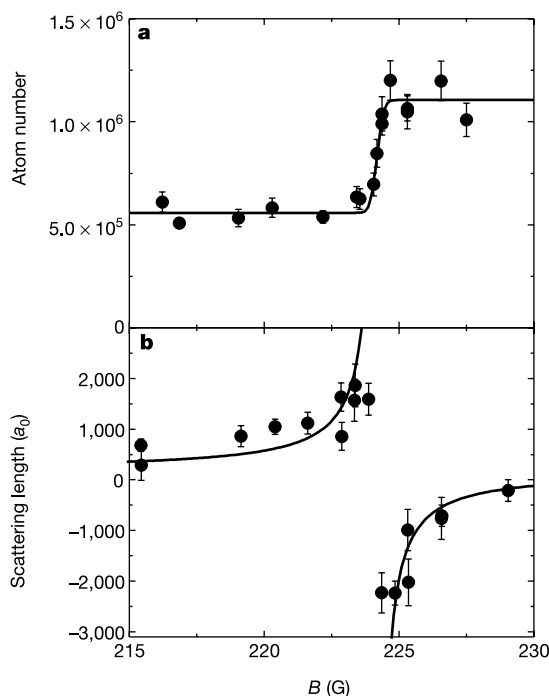


Figure 1 Atom loss following magnetic field ramps across a Feshbach resonance in ^{40}K . **a**, Number of $m_f = -5/2$ and $m_f = -9/2$ atoms as a function of the final magnetic field value B of a $40\text{ }\mu\text{s G}^{-1}$ ramp that starts at 227.81 G . The atom cloud is initially at $T/T_F = 0.21$ with $n_{\text{pk}} = 2.1 \times 10^{13}\text{ cm}^{-3}$, where n_{pk} is the peak atomic density of the two-component cloud. The relative number of $m_f = -9/2$ and $m_f = -5/2$ atoms remains approximately equal throughout the experiment. The data are fitted to an error function. The resulting gaussian width of the transition region is $0.21 \pm 0.07\text{ G}$, and the transition position is $B_0 = 224.18 \pm 0.05\text{ G}$. **b**, For comparison, we plot a previous measurement of the Feshbach resonance¹⁵. The resonance peak was found to be at $224.21 \pm 0.05\text{ G}$.

final magnetic field value of the ramp. We find that the atoms disappear abruptly at the Feshbach resonance peak.

We also investigate the atom loss as a function of the rate of the magnetic field sweep. Figure 2 illustrates the result of linear magnetic field ramps across the Feshbach resonance from 228.25 G to 216.15 G . For our fastest sweeps there is no observable effect upon the atoms, whereas significant atom number loss is observed for slower sweeps. For our slowest sweeps, we find that the number of atoms lost saturates at 50%. This loss cannot be explained by inelastic loss, because the atoms vanish at least two orders of magnitude more quickly than expected from previously measured inelastic collision rates at a resonance¹³. When we reverse the process by applying an additional magnetic field ramp across the Feshbach resonance in the opposite direction, we observe a return of nearly all the 'lost' atoms (Fig. 2). This is consistent with the lost atom number corresponding to trapped molecules. Further, we can measure the lifetime τ of the molecules by varying the time before conversion back to atoms. We find that $\tau \approx 1\text{ ms}$, presumably owing to vibrational quenching of these highly vibrationally excited molecules²².

Surprisingly, the number of molecules produced is very large, despite the fact that the Fermi gas is not described by a single wavefunction as is a BEC. In fact, the number of molecules is sufficiently large that the temperature of an initial atomic gas at $T/T_F = 0.13$ is three times lower than the molecular Bose-Einstein condensation temperature T_c in the optical trap. Our calculation of T_c assumes that these weakly bound molecules have a polarizability twice that of a single atom, and hence are subject to the same trapping frequency as the atoms²³. Future experiments could probe the distribution of energies in the ultracold molecular gas.

Using radio-frequency (r.f.) spectroscopy we directly probe the ultracold molecules. First, we create the molecules with a $40\text{ }\mu\text{s G}^{-1}$ magnetic field ramp. This ramp starts at 227.81 G , and ends at various final magnetic field values, B_{hold} , below the resonance. At B_{hold} , a $13\text{-}\mu\text{s}$ r.f. pulse is applied to the cloud; the radio frequency is

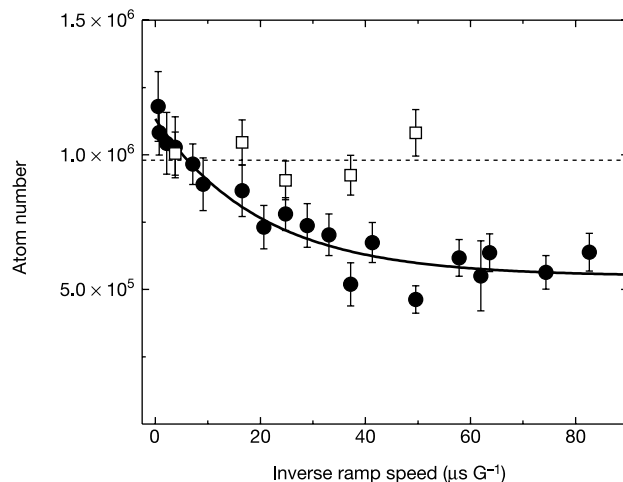


Figure 2 Dependence of atom loss on the magnetic field ramp rate through the Feshbach resonance. Filled circles, atom number following a magnetic field ramp across the Feshbach resonance from high to low field. The data are taken with $T/T_F = 0.33$ and $n_{\text{pk}} = 1.4 \times 10^{13}\text{ cm}^{-3}$. In accordance with a Landau-Zener model, we fit the data to an exponential (solid line); we find that the decay constant is $20 \pm 6\text{ }\mu\text{s G}^{-1}$. Data taken at $T/T_F = 0.13$ and $n_{\text{pk}} = 9 \times 10^{12}\text{ cm}^{-3}$ behave similarly. Open squares, data for which the magnetic field was first ramped downward at a rate of $40\text{ }\mu\text{s G}^{-1}$ across the resonance to create molecules, and then ramped back above the Feshbach resonance at varying rates to again create free atoms. The dashed line is the average of the square points.

chosen so that the photon energy is near the energy splitting between the $m_f = -5/2$ and $m_f = -7/2$ atom states. The resulting population in the $m_f = -7/2$ state, which is initially unoccupied, is then probed selectively either by separating the spin states spatially using a strong magnetic field gradient during free expansion (Stern–Gerlach imaging), or by leaving the magnetic field high (215 G) and taking advantage of nonlinear Zeeman shifts.

Figure 3a shows a sample r.f. spectrum at $B_{\text{hold}} = 222.49$ G; the resulting number of atoms in the $m_f = -7/2$ state is plotted as a function of the frequency of the r.f. pulse. We observe two distinct features in the spectrum. The sharp, symmetric peak is very near the expected $m_f = -5/2$ to $m_f = -7/2$ transition frequency for free atoms, ν_{atom} . With the Stern–Gerlach imaging we see that the total number of $m_f = -5/2$ and $m_f = -7/2$ atoms is constant, consistent with transfer between these two atom states. The width of this line is defined by the Fourier width of the applied r.f. pulse. Nearby is a broader, asymmetric peak shifted lower in frequency. Here we find that the total number of observed atoms ($m_f = -5/2 + m_f = -7/2$) after the r.f. pulse actually increases (Fig. 4). Also, the resulting $m_f = -7/2$ gas in this region has a significantly increased kinetic energy, which grows linearly for larger frequency shifts from the atom peak (Fig. 3b).

The asymmetric peak can be interpreted as the dissociation of molecules into free $m_f = -7/2$ and $m_f = -9/2$ atoms. Because the applied r.f. pulse stimulates a transition to a lower energy Zeeman

state, we expect $h\nu_{\text{rf}} = h\nu_{\text{atom}} - E_{\text{binding}} - \Delta E$, where h is Planck's constant, E_{binding} is the binding energy of the molecule, and we have ignored mean-field interaction energy shifts. The remaining energy, ΔE , must be imparted to the dissociating atom pair as kinetic energy. Further, the observed lineshape of the asymmetric peak should depend on the Franck–Condon factor, which gives the overlap of the molecular wavefunction with the wavefunction of the free atom pair.

We have calculated this multichannel Franck–Condon overlap as a function of energy. The resulting transition rate, convolved with the frequency width of the applied r.f. and scaled vertically, is shown as the solid line in Fig. 3a. The agreement between theory and experiment for the dissociation spectrum is quite good. This well-resolved spectrum provides much information about the molecular wavefunction. For example, the mean interatomic separation within each molecule at this magnetic field is extremely large, $170 a_0$, where a_0 is the Bohr radius.

In Fig. 5 we plot the magnetic field dependence of the frequency shift $\Delta\nu$ between the atom line and the threshold of the molecular spectrum, which should correspond to the molecular binding energy. Although this could in principle be obtained directly from the transfer spectrum (Fig. 3a), the appearance of the threshold in

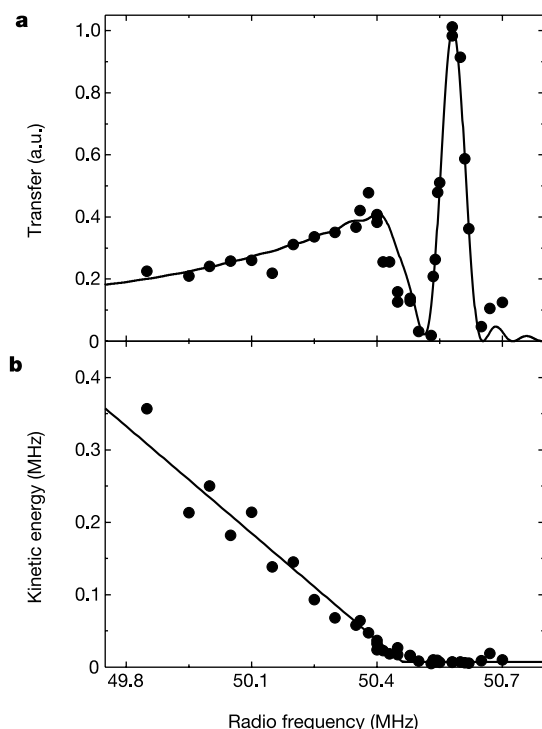


Figure 3 Photodissociation spectrum of ultracold molecules, and resulting energy per atom. **a**, Dissociation spectrum at $B_{\text{hold}} = 222.49$ G, with an original atom cloud containing $N = 1.4 \times 10^6$ atoms at $T/T_F = 0.26$, where N is the total number of atoms in the two-component cloud. The relative number of $m_f = -7/2$ atoms after an applied r.f. pulse is plotted versus radio frequency. The solid line is the weighted sum of the atomic transition line and the calculated dissociation spectrum, both convolved with the frequency width of the r.f. pulse. **b**, Resulting kinetic energy per $m_f = -7/2$ atom. Two separate linear fits are applied to the data to determine the threshold position. The slope beyond threshold is 0.49 ± 0.03 ; this indicates that the atom pair ($m_f = -7/2 + m_f = -9/2$) receives the additional energy ΔE beyond the binding energy when the molecule is dissociated.

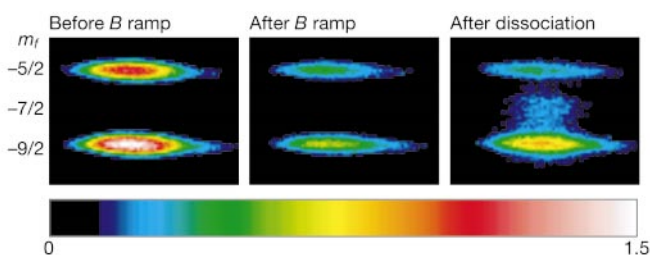


Figure 4 Absorption images of the quantum gas using a Stern–Gerlach technique. We start with ultracold fermionic atoms in the $m_f = -5/2$ and $m_f = -9/2$ states of ^{40}K . A magnetic field ramp through the Feshbach resonance causes 50% atom loss, owing to adiabatic conversion of atoms to diatomic molecules. To directly detect these bosonic molecules we apply an r.f. photodissociation pulse; the dissociated molecules then appear in the $m_f = -7/2$ and $m_f = -9/2$ atom states. The shaded bar indicates the optical depth.

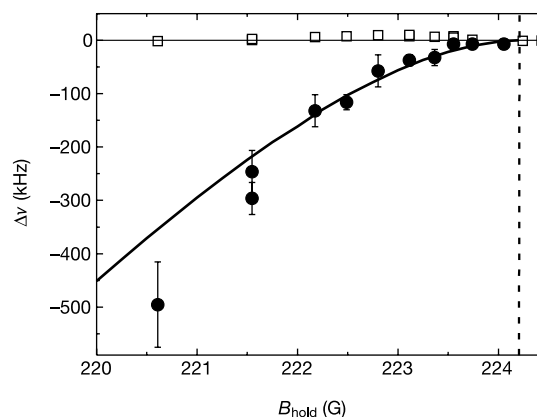


Figure 5 Binding energy of the molecules. The frequency shift $\Delta\nu$ from the expected $m_f = -5/2$ to $m_f = -7/2$ transition is plotted versus magnetic field for the atoms (open squares) and the molecules (filled circles). The typical atom cloud before molecule creation is characterized by $T/T_F = 0.14$ and $N = 7 \times 10^5$. The dashed line indicates the Feshbach resonance position. The solid line is a calculation of the binding energy of the molecules as a function of detuning from the resonance.

the energy of the $m_f = -7/2$ cloud is more clear (Fig. 3b). We compare the position of this energy threshold to the expected atom–atom transition frequency ν_{atom} based upon a calibration of the magnetic field strength. The data are consistent with a theoretical calculation of the binding energy (solid line) based upon a full coupled channels calculation with no free parameters.

This agreement with theory leaves no doubt that we are creating large numbers of weakly bound molecules. These highly vibrationally excited molecules could be used to study ultracold molecule–molecule, or molecule–atom, collisions^{22,24,25}. Further, the explicit coupling of a quantum degenerate gas of fermionic atoms to bosonic molecules could possibly be developed as a method to facilitate creation of a predicted exotic fermionic superfluid. In addition, our molecule detection technique could be extended to measure the gap energy in this superfluid phase^{26,27}. □

Received 29 April; accepted 19 May 2003; doi:10.1038/nature01738.

- Wynar, R., Freeland, R. S., Han, D. J., Ryu, C. & Heinzen, D. J. Molecules in a Bose–Einstein condensate. *Science* **287**, 1016–1019 (2002).
- Donley, E. A., Claussen, N. R., Thompson, S. T. & Wieman, C. E. Atom–molecule coherence in a Bose–Einstein condensate. *Nature* **417**, 529–533 (2002).
- Holland, M., Kokkelmans, S. J. J. M. F., Chiofalo, M. L. & Walser, R. Resonance superfluidity in a quantum degenerate Fermi gas. *Phys. Rev. Lett.* **87**, 120406 (2001).
- Timmermans, E., Furuya, K., Milloni, P. W. & Kerman, A. K. Prospect of creating a composite Fermi–Bose superfluid. *Phys. Lett. A* **285**, 228–233 (2001).
- Feshbach, H. A unified theory of nuclear reactions. II. *Ann. Phys. (NY)* **19**, 287–313 (1962).
- Stwalley, W. C. Stability of spin-aligned hydrogen at low temperatures and high magnetic fields: New field-dependent scattering resonances and predissociations. *Phys. Rev. Lett.* **37**, 1628–1631 (1976).
- Tiesinga, E., Verhaar, B. J. & Stoof, H. T. C. Threshold and resonance phenomena in ultracold ground-state collisions. *Phys. Rev. A* **47**, 4114–4122 (1993).
- Inouye, S. *et al.* Observation of Feshbach resonances in a Bose–Einstein condensate. *Nature* **392**, 151–154 (1998).
- Cornish, S. L., Claussen, N. R., Roberts, J. L., Cornell, E. A. & Wieman, C. E. Stable ⁸⁵Rb Bose–Einstein condensates with widely tunable interactions. *Phys. Rev. Lett.* **85**, 1795–1798 (2000).
- Loftus, T., Regal, C. A., Ticknor, C., Bohn, J. L. & Jin, D. S. Resonance control of elastic collisions in an optically trapped Fermi gas of atoms. *Phys. Rev. Lett.* **88**, 173201 (2002).
- Dieckmann, K. *et al.* Decay of an ultracold fermionic lithium gas near a Feshbach resonance. *Phys. Rev. Lett.* **89**, 203201 (2002).
- O'Hara, K. M. *et al.* Measurement of the zero crossing in a Feshbach resonance of fermionic ⁶Li. *Phys. Rev. A* **66**, 041401 (2002).
- Regal, C. A., Ticknor, C., Bohn, J. L. & Jin, D. S. Tuning p-wave interactions in an ultracold Fermi gas of atoms. *Phys. Rev. Lett.* **90**, 053201 (2003).
- O'Hara, K. M., Hemmer, S. L., Gehm, M. E., Granade, S. R. & Thomas, J. E. Observation of a strongly interacting degenerate Fermi gas of atoms. *Science* **298**, 2179–2182 (2002).
- Regal, C. A. & Jin, D. S. Measurement of positive and negative scattering lengths in a Fermi gas of atoms. *Phys. Rev. Lett.* (in the press).
- Bourdel, T. *et al.* Measurement of interactions energy near a Feshbach resonance in a ⁶Li Fermi gas. Preprint at (<http://arXiv.org/cond-mat/0303079>) (2003).
- Timmermans, E., Tommasini, P., Hussein, M. & Kerman, A. Feshbach resonances in atomic Bose–Einstein condensates. *Phys. Rep.* **315**, 199–230 (1999).
- Abeelen, F. A. & Verhaar, B. J. Time-dependent Feshbach resonance scattering and anomalous decay of a Na Bose–Einstein condensate. *Phys. Rev. Lett.* **83**, 1550–1553 (1999).
- Mies, F. H., Tiesinga, E. & Julienne, P. S. Manipulation of Feshbach resonance in ultracold atomic collisions using time-dependent magnetic fields. *Phys. Rev. A* **61**, 022721 (2000).
- Stenger, J. *et al.* Strongly enhanced inelastic collisions in a Bose–Einstein condensate near Feshbach resonances. *Phys. Rev. Lett.* **82**, 2422–2425 (1999).
- DeMarco, B. & Jin, D. S. Onset of Fermi degeneracy in a trapped atomic gas. *Science* **285**, 1703–1706 (1999).
- Soldán, P., Cvitas, M. T., Hutson, J. M., Honvault, P. & Launay, J.-M. Quantum dynamics of ultracold Na + Na₂ collisions. *Phys. Rev. Lett.* **89**, 153201 (2002).
- Ratcliff, L. B., Fish, J. L. & Konowalow, D. D. Electronic transition dipole moment functions for transitions among the twenty-six lowest-lying states of Li₂. *J. Mol. Spectrosc.* **122**, 293–312 (1987).
- Balakrishnan, N., Forrey, R. C. & Dalgarno, A. Quenching of H₂ vibrations in ultracold ³He and ⁴He collisions. *Phys. Rev. Lett.* **80**, 3224–3227 (1998).
- Forrey, R. C., Balakrishnan, N., Dalgarno, A., Haggerty, M. R. & Heller, E. J. Quasiresonant energy transfer in ultracold atom–diatom collisions. *Phys. Rev. Lett.* **82**, 2657–2660 (1999).
- Petrosyan, K. G. Fermionic atom laser. *JETP Lett.* **70**, 11–16 (1999).
- Torma, P. & Zoller, P. Laser probing of atomic Cooper pairs. *Phys. Rev. Lett.* **85**, 487–490 (2000).

Acknowledgements We thank E. A. Cornell, C. E. Wieman, C. H. Greene and S. Inouye for discussions. This work was supported by the NSF and NIST; C.A.R. acknowledges support from the Hertz Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.A.R. (regal@jilau1.colorado.edu).

Achromatic Fresnel optics for wideband extreme-ultraviolet and X-ray imaging

Yuxin Wang*, Wenbing Yun* & Chris Jacobsen*†

* Xradia, Inc., 4075A Sprig Drive, Concord, California 94520, USA

† Department of Physics & Astronomy, Stony Brook University, Stony Brook, New York 11794, USA

Advances in extreme-ultraviolet (EUV) and X-ray optics are providing powerful new capabilities in high-resolution imaging and trace-element analysis of microscopic specimens¹, and the potential for fabricating devices of smaller critical dimensions in next-generation integrated circuit lithography². However, achieving the highest resolution with such optics usually requires the illuminating EUV or X-ray beam to be highly monochromatic. It would therefore be highly desirable to have large-field-of-view, sub-100-nm resolution optics that are achromatic to a significant degree, allowing more light to be utilized from broader bandwidth sources such as laser-produced plasmas. Here we report an achromatic Fresnel optical system for EUV or X-ray radiation that combines a Fresnel zone plate with a refractive lens with opposite chromatic aberration. We use the large anomalous dispersion property of the refractive lens material near an absorption edge to make its fabrication practical. The resulting structure can deliver a resolution comparable to that of the Fresnel zone plates that have achieved the highest resolution (25 nm; ref. 3) in the entire electromagnetic spectrum, but with an improvement of two or more orders of magnitude in spectral bandwidth.

When Röntgen discovered X-rays in 1895, he immediately searched for a means to focus them^{4,5}. On the basis of uncertain observations of slight refraction in prisms, he stated that the index of refraction for X-rays in materials must be no more than about $n = 1.05$ if it differed at all from unity. Two decades later, Einstein⁶ proposed that the refractive index for X-rays was $n = 1 - \delta$ with $\delta \approx 10^{-6}$, allowing for total external reflection at grazing incidence angles satisfying $\cos \theta_c = 1 - \delta$, giving $\theta_c \approx \sqrt{2\delta}$ or of the order of 1 mrad (here θ_c is the critical angle). Modern measurements give critical angles of, for example, 9 mrad for 8.98-keV X-rays from gold surfaces; nonetheless, this sets a fundamental limit to the maximum numerical aperture, NA (and thus spatial resolution), of polychromatic X-ray focusing optics. While these optics have advanced from early demonstrations⁷ to sub-500-nm focusing⁸, demonstrated routes to higher spatial resolution EUV and X-ray imaging and lithography have instead involved optics for monochromatic radiation. Such optics include Fresnel zone plates⁹ with 25-nm potential image resolution³ and high efficiency; simple¹⁰, Fresnel¹¹, and compound¹² refractive lenses with spatial resolution to about 200 nm (ref. 13); and mirror optics coated with synthetic multilayers¹⁴ which have achieved 50-nm resolution in lithography applications¹⁵. These approaches have led to significant advances in EUV and X-ray science.

However, all of these higher-resolution methods require the illuminating EUV or X-ray beam to be highly monochromatic. For certain experiments, such as absorption spectroscopy with focused, monochromatized X-ray beams, the demands of these optics for narrow spectral bandwidth illumination are easily met without compromising experimental throughput. With broader-bandwidth EUV or X-ray sources, such as laser-produced plasmas and line emission from X-ray targets, the availability of achromatic optics would allow significantly more light to be utilized for experiments and applications.

The achromatic Fresnel objective combines a high-spatial-resolution Fresnel zone plate with a refractive lens operating in the anomalous dispersion condition to achieve these desirable properties. We begin by considering the properties of Fresnel zone plates. These optics make use of a series of concentric absorbing or phase-shifting rings with prescribed radii, giving a focal length of

$$f_Z = \frac{R^2}{\lambda N} = \frac{2R\Delta R}{\lambda} \quad (1)$$

where $2R$ is the diameter of the optic, ΔR is the width of the finest, outermost zone of zone number N , and λ is the wavelength. For imaging at the Rayleigh limit of $0.61\lambda/\text{NA}$, the spatial resolution is $1.22\Delta R$. The number of zones N determines the required illumination monochromaticity or spectral bandwidth:

$$\frac{N}{2} = \frac{R}{4\Delta R} = \frac{\lambda}{\Delta\lambda} \quad (2)$$

Combining these requirements, we see that if we wish to have a large diameter $2R$ for a large field of view and long focal length, large values of N and thus narrow bandpass $\Delta\lambda/\lambda = 2/N$ radiation is required. For example, a zone plate with a diameter of a few millimetres and $<100\text{-nm}$ outer zone width would have over

10,000 zones, and therefore a frequently impractical 0.01% bandwidth.

We now consider refractive lenses. For a thin lens with a single curved surface of radius R_C , the focal length is given by

$$f_R = \frac{R_C}{n-1} \quad (3)$$

where n is the refractive index of the material. At wavelengths shorter than about 50 nm, the refractive index has the form

$$n = 1 - (\delta + i\beta) = 1 - \alpha\lambda^2(f_1 + if_2) \quad (4)$$

where $\alpha = n_a r_e / 2\pi$. In this expression, n_a is the number of atoms per volume, $r_e = 2.8 \times 10^{-15} \text{ m}$ is the classical radius of the electron, and $(f_1 + if_2)$ is the complex number of effective electrons per atom¹⁶. Ignoring the absorptive part f_2 of the refractive index, we see that the focal length of a refractive lens for EUV and X-ray radiation is given by

$$f_R = \frac{R_C}{-\alpha\lambda^2 f_1} \quad (5)$$

so that a concave lens produces positive focusing when $f_1 > 0$, the opposite of the usual case for visible light.

A refractive lens is strongly chromatic, as its focal length scales as the inverse of λ^2 and $f_1(\lambda)$. Moreover, its focal length is too weak to be used as a single focusing optic because the quantity $\alpha\lambda^2 f_1$ has the value of 10^{-3} to 10^{-6} for EUV light and X-rays, as anticipated by Einstein. It is only with 'hard' X-rays of energy greater than about 5 keV, where absorption is less problematic, that multiple weak refractive lenses can be stacked up into a compound refractive X-ray lens¹². The image resolution at present achieved is in the 300-nm range, being limited by absorption, imperfections in the fabrication of the refracting thickness profile, and alignment and depth of focus effects¹³.

The achromatic Fresnel optic combines these two optics to cancel the chromatic aberration over a reasonable wavelength range (Fig. 1; other combinations of diffractive and refractive optics have been used to obtain different chromatic effects¹⁷, and zone plate achromatic doublets have also been proposed¹⁸). Let us consider the imaging properties of each lens as the wavelength is varied from a reference value λ as $\lambda + \Delta\lambda$. The focal length of a Fresnel zone plate is then given by:

$$f_Z' \approx f_Z \frac{1}{1 + \frac{\Delta\lambda}{\lambda}} \quad (6)$$

With the refractive lens, we must consider the variation both in $\lambda^2 \rightarrow \lambda^2 + 2\lambda\Delta\lambda + (\Delta\lambda)^2$, and also in f_1 . It is well known that the real part of the refractive index f_1 can have a strong dependence on wavelength near an absorption resonance owing to anomalous dispersion. This effect is also exploited in multiple-wavelength anomalous dispersion (MAD) methods in protein crystallography¹⁹. The simplest approximation that we can make to f_1 is to incorporate a linear dispersive term and write it as:

$$f_1 \rightarrow f_1 + \frac{\Delta f_1}{\Delta\lambda} \Delta\lambda \quad (7)$$

Ignoring terms of order $[\Delta\lambda/\lambda]^2$, the focal length of a refractive lens becomes:

$$f_R' \approx f_R \frac{1}{1 + 2\frac{\Delta\lambda}{\lambda} + \frac{\Delta f_1}{f_1}} \quad (8)$$

If the two lenses are placed in close proximity to each other, their reciprocal focal lengths can be added:

$$\frac{1}{f_C} = \frac{1}{f_Z'} + \frac{1}{f_R'} = \frac{1}{f_Z} + \frac{1}{f_R} + \frac{\Delta\lambda}{\lambda} \left[\frac{1}{f_Z} + \frac{1}{f_R} (2 + D) \right] \quad (9)$$

where $D \equiv (\Delta f_1/f_1)/(\Delta\lambda/\lambda)$ characterizes the dispersion. If the term in the bracket becomes zero, the focal length becomes

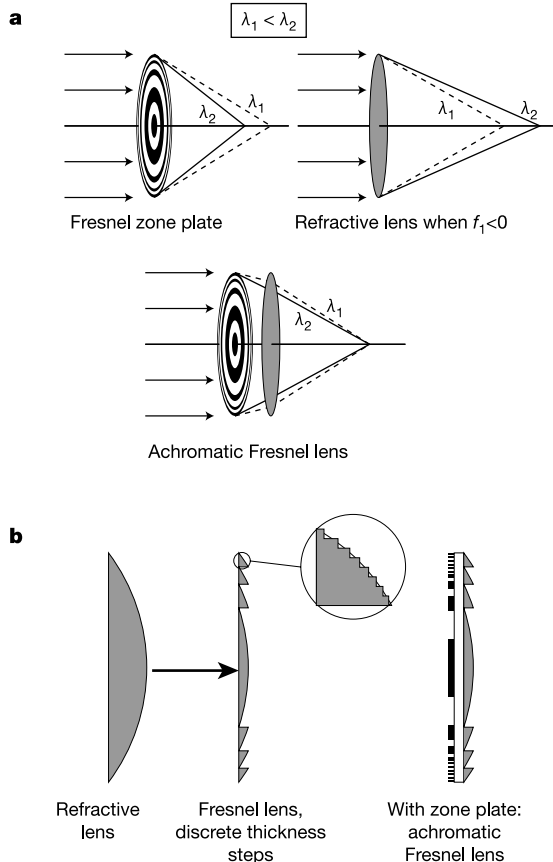


Figure 1 The principle of an achromatic Fresnel lens, and an example of how it can be realized in a single optic. The chromatic aberrations of a zone plate produce a longer focal length for shorter wavelengths (λ_1) than for longer wavelengths (λ_2) (a). The chromatic aberrations of a refractive lens are opposite; a shows a convex refractive lens for wavelength regions where f_1 is decreasing as the wavelength λ is decreased. One can thus combine the two optic types to cancel chromatic aberration over a certain wavelength range. The simplest such combination (b) involves fabrication of a zone plate and a plano-convex refractive lens on opposite sides of a single window; one can also replace the refractive lens with a refractive Fresnel lens to minimize absorption, and also use a staircase approximation to the curvature of the refractive Fresnel lens surface.

independent of $\Delta\lambda$; that is, we have an achromat. Note that one can also seek large values of the bracketed term if it is desired to enhance the chromatic dispersion of a zone plate. The achromat condition can be written as:

$$\frac{f_R}{f_Z} = -(2 + D) \quad (10)$$

The values of f_1 and D can be calculated from absorption data by the Kramers–Kronig relations^{16,20}, such as are shown in Fig. 2 for the Si and Cu L edges. Near the absorption edges, the magnitude of D can be significantly larger than 2, so that the focal length of the refractive lens is much longer than that of the zone plate (several metres versus several centimetres in our examples). As a result, the refractive lens by itself contributes only weakly to the net focusing action (it is only used as a dispersive corrector), so that the spatial resolution is determined almost solely by the Fresnel zone plate. This long refractive focal length goes along with a larger, easier-to-fabricate radius of curvature for the refractive lens component; this radius of curvature R_C can be written as:

$$R_C \approx \alpha(2R\Delta R)\lambda^2 \frac{\Delta f_1}{\Delta\lambda} \quad (11)$$

Since $(\Delta f_1/\Delta\lambda)$ tends to be largest when it has a positive value near X-ray absorption edges (see Fig. 2), convex or plano-convex solutions for the refractive correction lens are preferred. In addition,

the use of wavelengths longer than the absorption edge is preferred for EUV light and soft X-rays to reduce absorption. Finally, we note that the bandwidth gain depends on both the number of zones in the zone plate and on the dispersion properties of the refractive lens material. A schematic illustration of the effect is shown in Fig. 1.

The use of anomalous dispersion increases the radius of curvature and thus minimizes the total thickness of convex refractive lens components. However, at EUV wavelengths in particular the absorption from the thicker lens regions may be unacceptable. One solution is to replace the refractive lens with a refractive Fresnel lens so that the overall curvature can be maintained within a stepwise approximation. Ideally these thickness steps shift the phase of the transmitted light by an integer multiple of 2π . Other thickness steps can be chosen if the resulting phase error is compensated by adjustment of the zone positions of the Fresnel zone plate. It should be noted that an aberration-free Fresnel lens imposes a certain requirement for monochromaticity. As indicated by equation (4), the accumulated phase difference between a wave that propagates a distance t through free space and one that propagates the same distance through a material is $kt\delta$, so that a wavefront is displaced by a distance of $t\delta$. For illumination with a bandwidth of $\Delta\lambda$, the coherence length is given by $\lambda^2/(\Delta\lambda)$, so one must have $t\delta < \lambda^2/(\Delta\lambda)$ or $t < (\lambda/\Delta\lambda)\lambda/\delta$. If we consider the example of Fig. 2 of $\lambda = 13.5$ nm in Si where $\delta = 9.5 \times 10^{-4}$ and

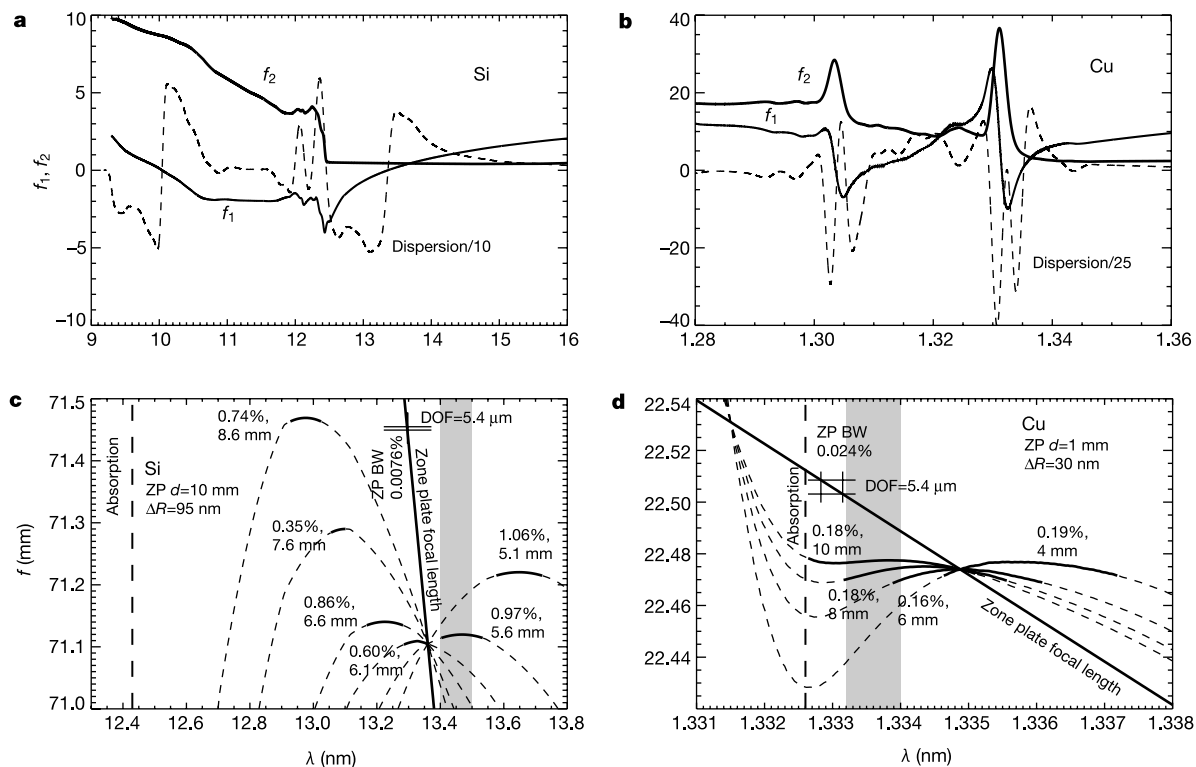


Figure 2 X-ray optical constants and their consequences for EUV and soft X-ray achromatic Fresnel objectives. At top (**a**, **b**) are shown the oscillator strengths ($f_1 + if_2$) and the dispersion D for silicon and copper L edges. The absorptive term f_2 was determined from thin-film absorption measurements Fourier-filtered with a Hanning cut-off of 0.2 eV (Si; from ref. 25) or 0.7 eV (Cu in SrCuO₂; from ref. 26). The phase-shifting term f_1 was calculated by inserting the measured near-absorption-edge data into tabulated values for f_2 , and then using this data to modify tabulated values of f_1 using the Kramers–Kronig relationship¹⁶; this procedure was able to reproduce experimental measurements²⁷ of near-edge f_1 values from near-edge f_2 data of amorphous carbon. At bottom (**c**, **d**) are shown examples of achromatic Fresnel objective configurations at wavelengths longer than the absorption edge of each element. The focal length (f) of a

Fresnel zone plate (ZP) changes significantly over the wavelength range shown, especially when compared with the depth of field $2\lambda/(\text{NA})^2$ of the optic, so that only very narrow bandwidth (BW) radiation can be used for imaging. However, when an achromatic Fresnel lens is produced by combining the zone plate with a refractive lens, the combined focal length (for several values of refractive lens radius of curvature R_C) is constant to within a depth of field, DOF (as indicated by thickened lines for the combined focal length), over bandwidths approaching 1% for Si and 0.2% for Cu with these sets of parameters. In **c**, we indicate with a grey band the 13.4–13.5 nm wavelength band used in EUV lithography systems, while in **d** the grey band corresponds to the Cu L α fluorescence line. Note that improved measurements of absorption spectra and thus f_2 values would lead to refined estimates of achromatic bandwidth. d , diameter of the zone plate.

the allowable achromatic bandwidth is 0.97% or $(\lambda/\Delta\lambda) = 103$, we see that a thickness difference of up to 1.5 mm between the refractive lens and the refractive Fresnel lens profile still satisfies the coherence length requirement for the bandwidth used. It is this net thickness difference that is the important parameter; one could go up to this limit in $(\lambda/\Delta\lambda) = 103$ steps of 2π phase change, or 34 steps of 6π phase change, to pick two examples. Finally, since the radius r of a plano-convex lens with a thickness t at its centre is given by $r = \sqrt{2R_C t - t^2}$, with $R_C = 5.6$ mm and $t = 1.5$ mm one can have a lens radius up to 3.8 mm (or a numerical aperture of up to 0.75) and still remain within the coherence length requirement of a Fresnel lens with 0.93% bandwidth.

In practice, one can imagine lithographically fabricating an achromatic Fresnel optic (AFO) on a single thin membrane, with the Fresnel zone plate fabricated on one side, and a refractive Fresnel lens on the other side, as shown in Fig. 1 (possibly fabricated using multilevel^{21,22} or imprint²³ methods). The achievable efficiencies of the zone plate and the Fresnel lens are about 30–50% and 30–80%, respectively, leading to a combined efficiency of 10–40% for the AFO which compares quite favourably with, for example, the ~5% net throughput of a next-generation lithography system using six optics with 60% reflectivity each to gain the required resolution and field of view.

The size of the AFO and its imaging field are likely to be limited by primary aberrations. Seidel wavefront aberrations for imaging finite conjugates with zone plate optics have been calculated by Young²⁴. At 1.34-nm wavelength (Cu L absorption edge) with a 4:1 demagnifying geometry (a standard set-up used in lithography cameras), aberration-free imaging fields of between 2 and 15 mm can be expected for outermost zone widths between 40 and 95 nm without aberration correction. When used with EUV radiation near 12.5 nm wavelength, the NAs are increased nearly tenfold and primary aberrations become more problematic unless the field of view is kept to a few tenths of 1 mm. Spherical aberration does not exist for Fresnel zone plates when the zone placement is computed for a specific imaging geometry. This property, along with the achromatic nature of the AFO, allows one to reduce the field curvature and astigmatism to an acceptable level by increasing the AFO diameter (therefore the focal length) while maintaining the same field of view. Coma can be reduced and eliminated in some cases by appropriate aperturing.

In general, the AFO design provides two important benefits: it makes a large bandwidth of electromagnetic radiation usable, and it allows large-diameter high-resolution optics to be produced without suffering from chromatic aberration. The increased bandwidth should greatly improve the throughput of applications using broadband X-ray sources, such as X-ray tubes and laser-produced plasmas. In addition, non-spectroscopic imaging applications such as tomography at synchrotron radiation sources could operate with reduced imaging time using multilayer monochromators rather than narrow-bandpass crystal monochromators. The large diameter gives large working distance and large imaging field, both valuable attributes in imaging applications. The present AFO design has the potential to make significant impact in X-ray microscopy and microanalysis, and next-generation lithography applications. □

Received 27 January; accepted 19 May 2003; doi:10.1038/nature01756.

1. Meyer-Ilse, W., Warwick, T. & Attwood, D. (eds) *X-ray Microscopy: Proceedings of the Sixth International Conference* (American Institute of Physics, Melville, New York, 2000).
2. Chapman, H. N. *et al.* First lithographic results from the extreme ultraviolet engineering test stand. *J. Vac. Sci. Technol. B* **19**, 2389–2395 (2001).
3. Peuker, M. High-efficiency nickel phase zone plates with 20 nm minimum outermost zone width. *Appl. Phys. Lett.* **78**, 2208–2210 (2001).
4. Röntgen, W. C. Über eine neue Art von Strahlen. Vorläufige Mittheilung. *Sber. Phys.-Med. Ges. Würzb.* **137**, 132–141 (1895).
5. Röntgen, W. C. On a new kind of rays. *Nature* **53**, 274–276 (1896).
6. Einstein, A. Lassen sich Brechungsexponenten der Körper für Röntgenstrahlen experimentell ermitteln? *Verh. Dtsch. Phys. Ges.* **20**, 86–87 (1918).

7. Kirkpatrick, P. & Baez, A. V. Formation of optical images by x-rays. *J. Opt. Soc. Am.* **38**, 766–774 (1948).
8. Hignette, O. *et al.* in *X-ray Micro- and Nano-Focusing: Applications and Techniques II* (ed. McNulty, I.) 105–116 (Proc. SPIE, The International Society for Optical Engineering, San Diego, 2001).
9. Baez, A. V. A self-supporting metal Fresnel zone-plate to focus extreme ultra-violet and soft X-rays. *Nature* **186**, 958 (1960).
10. Michette, A. G. *Optical Systems for Soft X-Rays* (Plenum, New York, 1986).
11. Yang, B. X. Fresnel and refractive lenses for X-rays. *Nucl. Instrum. Meth. Phys. Res. A* **328**, 578–587 (1993).
12. Snigirev, A., Kohn, V., Snigireva, I. & Lengeler, B. A compound refractive lens for focusing high energy X-rays. *Nature* **384**, 49–51 (1996).
13. Lengeler, B. *et al.* Parabolic refractive X-ray lenses. *J. Synchrotron Radiat.* **9**, 119–124 (2002).
14. Spiller, E. Low-loss reflection coatings using absorbing materials. *Appl. Phys. Lett.* **20**, 365–367 (1972).
15. Naulleau, P. *et al.* Sub-70-nm EUV lithography at the Advanced Light Source static microfield exposure station using the ETS Set-2 optic. *J. Vac. Sci. Technol. B* **20**, 2829–2833 (2002).
16. Henke, B. L., Gullikson, E. M. & Davis, J. C. X-ray interactions: photoabsorption, scattering, transmission, and reflection at $E = 50$ –30,000 eV, $Z = 1$ –92. *At. Data Nucl. Data Tables* **54**, 181–342 (1993).
17. Dobson, S. L., Sun, P.-C. & Fainman, Y. Diffractive lenses for chromatic confocal imaging. *Appl. Opt.* **36**, 4744–4748 (1997).
18. Michette, A. G., Buckley, C., Gallo, F., Powell, K. & Pfauntsch, S. J. in *Advances in X-ray Optics* (ed. Freund, A. K. *et al.*) 303–310 (Proceedings SPIE Vol. 4145, The International Society for Optical Engineering, Bellingham, Washington, 2000).
19. Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
20. Jackson, J. D. *Classical Electrodynamics* (Wiley & Sons, New York, 1975).
21. Stern, M. B. & Medeiros, S. S. Deep three-dimensional microstructure fabrication for infrared binary optics. *J. Vac. Sci. Technol. B* **10**, 2520–2525 (1992).
22. Walsby, E. D. *et al.* Multilevel silicon diffractive optics for terahertz waves. *J. Vac. Sci. Technol. B* **20**, 2780–2783 (2002).
23. Hirai, Y. *et al.* Imprint lithography for curved cross-sectional structure using replicated Ni mold. *J. Vac. Sci. Technol. B* **20**, 2867–2871 (2002).
24. Young, M. Zone plates and their aberrations. *J. Opt. Soc. Am.* **62**, 972–976 (1972).
25. van Buuren, T. W. H. *et al.* Electronic structure of silicon nanocrystals as a function of particle size. *Abstr. Pap. Am. Chem. Soc.* **213**, 313 (1997).
26. Liu, R. S. *et al.* Evidence for electron-doped (*n*-type) superconductivity in the infinite-layer $(\text{Sr}_{0.9}\text{La}_{0.1})\text{CuO}_2$ compound by X-ray absorption near-edge spectroscopy. *Solid State Commun.* **118**, 367–370 (2001).
27. Dambach, S. *et al.* Novel interferometer in the soft x-ray region. *Phys. Rev. Lett.* **80**, 5473–5476 (1998).

Acknowledgements We thank J. Kirz for discussions, and S. Frigo for information about the calculation of Kramers–Kronig transforms.

Competing interests statement The authors declare competing financial interests: see the website (<http://www.nature.com>) for details.

Correspondence and requests for materials should be addressed to W.Y. (wyun@xradia.com).

Asymmetric pores in a silicon membrane acting as massively parallel brownian ratchets

Sven Matthias & Frank Müller

Max Planck Institute of Microstructure Physics, Weinberg 2, Halle 06120, Germany

The brownian motion of mesoscopic particles is ubiquitous and usually random. But in systems with periodic asymmetric barriers to movement, directed or ‘rectified’ motion can arise and may even modulate some biological processes¹. In man-made devices, brownian ratchets and variants based on optical or quantum effects have been exploited to induce directed motion^{2–14}, and the dependence of the amplitude of motion on particle size has led to the size-dependent separation of biomolecules^{6,8,15}. Here we demonstrate that the one-dimensional pores of a macroporous silicon membrane¹⁶, etched to exhibit a periodic asymmetric variation in pore diameter, can act as massively parallel and multiply stacked brownian ratchets that are potentially suitable for large-scale particle separations. We

show that applying a periodic pressure profile with a mean value of zero to a basin separated by such a membrane induces a periodic flow of water and suspended particles through the pores, resulting in a net motion of the particles from one side of the membrane to the other without moving the liquid itself. We find that the experimentally observed pressure dependence of the particle transport, including an inversion of the transport direction, agrees with calculations^{17,18} of the transport properties in the type of ratchet devices used here.

Particle separation is often accomplished by exploiting the dependence of drift or diffusion on size. That is, the process starts with a mixture, and as the particles move with different velocities, different components can be picked up at different positions away from the starting point. In contrast, ratchet-based devices¹⁴ tend to exploit changes in drift direction as a function of particle size, which can in principle lead to smaller devices capable of continuous operation. Microscopic calculations of the transport of particles in ratchet devices similar to those used here have revealed the closely related effects of the dependence of the net motion direction of the particles on particle size (for a fixed pressure) and on pressure amplitude (for fixed particle size)¹⁷. In these experiments, we demonstrate that the transport in our membranes exhibits the predicted pressure dependence.

Our experimental set-up consists of two basins filled with an aqueous solution of dispersed particles and separated by a horizontally mounted ratchet membrane (Fig. 1). The upper basin is open, and the lower basin closed and connected to an electrically driven pressure oscillator. The pore arrays were grown by a photo-electrochemical etching process^{19,20}, which has been extended to allow very strong diameter modulations^{16,21,22}. The particles were commercial luminescent polystyrene spheres with well-defined diameters of 0.1 μm , 0.32 μm , 0.53 μm and 1 μm , respectively. The concentration used was approximately one particle per pore. Interactions between the beads can therefore be neglected. The total number of particles in the upper basin was measured by photoluminescence (PL).

Applying square wave pressure oscillations to the lower basin generated a periodic back and forth streaming of the water-plus-particle system between the two basins through the membrane. By simultaneously recording the pressure in the lower basin and the filling level in the upper basin, we could investigate the water flow.

For carefully prepared hydrophilic pore walls and hyper-pure water as carrier liquid, the well-known macroscopic hydrodynamic equations are appropriate. The flow through the straight pores can be described by the Hagen–Poiseuille law. For the shaped pores the flow was still well approximated by the Hagen–Poiseuille law when using the minimum diameter. In addition, the oscillations of the particles together with the carrying water through the capillaries was directly observed by lock-in PL-spectroscopy using the pressure oscillations as reference signal. We modulated our whole experiment and toggled the driving pressure oscillations on and off every 60 seconds to exclude long-term drifts caused, for example, by water evaporation. This toggling also enables us to obtain more information about the system. Together with the macroscopic simulations presented later, this helps us to understand in more detail the drift and diffusion contributions.

In Fig. 2a the curve labelled U shows the evolution of the PL signal, starting from a homogeneous particle distribution. The applied root-mean-square (r.m.s.) pressure during the ‘on’ phase is 2,000 Pa with a frequency of 40 Hz. The ‘on’ phases of the cycles are highlighted in grey. When switched on for the first time at $t = 60$ s, the PL signal increases nearly linearly with time. After 60 s the total PL intensity and therefore the total number of particles in the upper basin has increased by about 8%. Turning off the periodic pressure oscillations leads to a nearly constant signal. Only a very slight decrease is visible. Repeating this cycle several times, we observe an increase of the signal during the ‘on’ phases of the cycles, although for longer times the number of added particles is reduced. On the other hand, during the ‘off’ phases the number of vanishing particles strongly increases with time. The profile changes from a stepwise function to a sawtooth-like behaviour and suggests the onset of a mechanism that pushes the particles back into the membrane.

Our experimental set-up does not permit us to measure the PL in the lower basin. In a second experiment we therefore mounted the membrane with reversed orientation. In Fig. 2a the curve labelled U_{reversed} shows the corresponding PL evolution. There is no pronounced systematic correlation with the ‘on’ and ‘off’ phases to be seen, but the number of particles is, except for some perturbations, clearly decreasing with time. A second control experiment was performed using straight cylindrical pores with a diameter of 2.4 μm . Curve $U_{\text{cylindrical}}$ in Fig. 2a shows that particle transport is

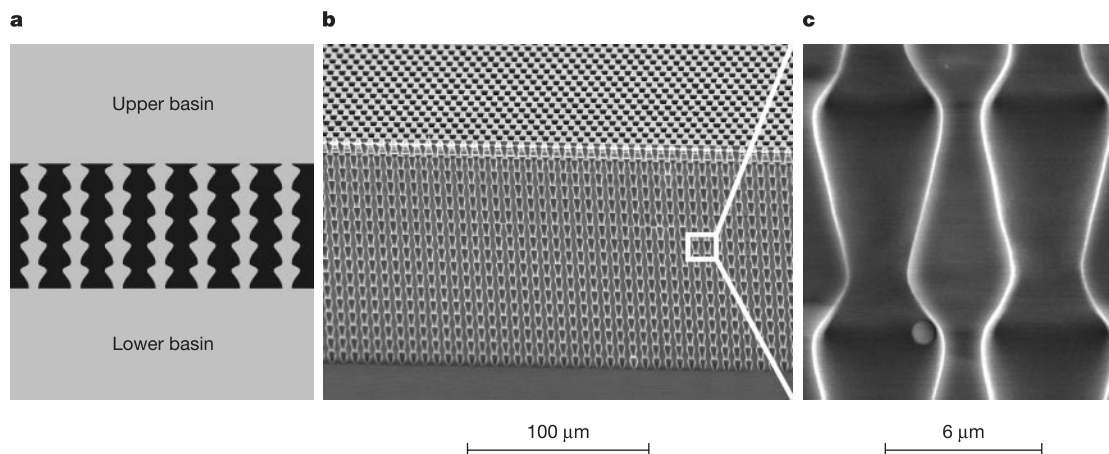


Figure 1 Experimental set-up. **a**, Schematic representation. The horizontally mounted membrane structure with asymmetric diameter modulated pores separates two basins, upper and lower. **b**, Scanning electron micrograph of a cleaved modulated macroporous silicon wafer. The macropores are arranged in a triangular lattice with a pitch of 6 μm and

a depth of about 150 μm . **c**, Scanning electron micrograph of a cleaved modulated macroporous silicon ratchet membrane. After drying, some colloidal spheres with a diameter of 1 μm stick to the silicon surface. The maximum pore diameter is 4.8 μm and the minimum 2.5 μm . The length of one period is 8.4 μm .

indeed only observable for the shaped pores.

Two processes are superimposed during the water-plus-particle oscillations: the brownian motion of the particles and its oscillations through the channel itself. The particle is diffusing randomly between liquid layers of different speed. Owing to the spatial asymmetry the liquid velocity field, and hence the friction forces onto the beads, possess a ratchet-like profile, enabling uni-directional motion^{12,17}. At present there are no convincing 'hand-waving' arguments for why the direction of transport depends on the amplitude of the applied pressure, as predicted by the numerical calculations.

The transport was microscopically modelled¹⁷ by solving the equation of motion for a single particle in an infinitely long pore given by the Langevin equation:

$$\dot{\mathbf{x}}(t) = \mathbf{v}(\mathbf{x}(t), t) + \sqrt{2D_0}\xi(t) \quad (1)$$

Here D_0 is the diffusion coefficient, $\xi(t)$ the independent gaussian noise and $\mathbf{v}(\mathbf{x}(t), t)$ the velocity field of the liquid. The microscopic calculations result in a systematic particle drift with a drift velocity v_m in the range of $\pm 1 \mu\text{m s}^{-1}$, depending on particle size and pressure amplitude. In addition, the calculations show that under these conditions the spreading of a particle ensemble can be described by an enhanced diffusion coefficient of the particles

$D_{\text{eff}} = rD_0$. Typically the enhancement factor r is in the range 1 to 10.

To understand the experimental data of Fig. 2a, we modelled the macroscopic system by a one-dimensional diffusion equation for the particle density $n(z, t)$.

$$\frac{\partial n}{\partial t} = -\frac{\partial}{\partial z}(j_{\text{drift}} + j_{\text{diffusion}}) = -\frac{\partial}{\partial z}\left(nv_m - \frac{\partial}{\partial z}(Dn)\right) \quad (2)$$

Typical parameters were taken from microscopic theory¹⁷. A transition region of $20 \mu\text{m}$ was introduced between membrane and basin where the parameters evolve gradually to reflect the smoothing caused by the oscillations.

The total number of particles in the upper and lower basins is shown in Fig. 2b. Comparing the results of the simulation for the upper basin with the experimental behaviour in Fig. 2a (U) we observe excellent qualitative agreement. All the characteristic peculiarities are reproduced. The signals show the same behaviour going from a stepwise function to a more sawtooth-like profile. After about 700 s some sort of saturation is observed where the net increase of particles over one cycle is strongly reduced compared to the beginning of the experiment. The evolution for the lower basin in the simulation again qualitatively reproduces the smooth, unstructured signal of the experiment with reversed mounted membrane (U_{reversed}).

Looking at the position dependence of the concentration profile (Fig. 3), we can immediately understand the reason for the observed evolutions. During the 'on' phases of the cycles the particles drift from the membrane to the upper basin. Since outside the membrane there is no drift, the particles accumulate near the surface. A large gradient towards the membrane evolves over time. This leads to the strong back-diffusion that is visible as soon as the drift is shut off. After 720 s almost the whole membrane is depleted. The net transport capacity of the membrane is limited by the few particles that diffuse into the membrane from the lower basins. The particles from the upper basin mostly diffuse back into the surface region of the membrane during the 'off' phase, only to be driven out again during the 'on' phase. Thus the characteristic triangular profile evolves. The low diffusion current into the volume of the upper basin and the limited diffusion current of particles from the lower

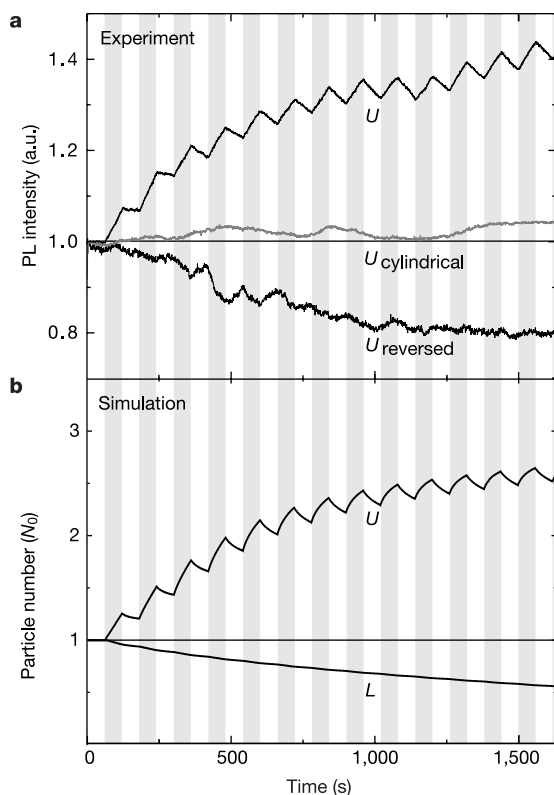


Figure 2 Photoluminescence intensity as a function of time. The pressure oscillations are toggled on and off every 60 s. The 'on' phases of the cycles are highlighted in grey.

a, Measured intensity profile in the upper basin (U) and in a second experiment for the reversed mounted membrane of the former lower basin (U_{reversed}) as well as for cylindrical pores ($U_{\text{cylindrical}}$) for an applied root mean square (r.m.s.) pressure during the 'on' phase of 2,000 Pa, an oscillation frequency of 40 Hz and a particle diameter of $0.32 \mu\text{m}$. **b**, Simulated particle number in the two basins as a function of time. The total number of particles N in units of the initial particle number N_0 in the upper (U) and lower (L) basins were obtained by numerical integration of equation (2) using typical parameter values from ref. 17: $v_m = 0.5 \mu\text{m s}^{-1}$, $D_0 = 1.5 \mu\text{m}^2 \text{s}^{-1}$, $r = 3$, membrane thickness $150 \mu\text{m}$, basin thickness $100 \mu\text{m}$.

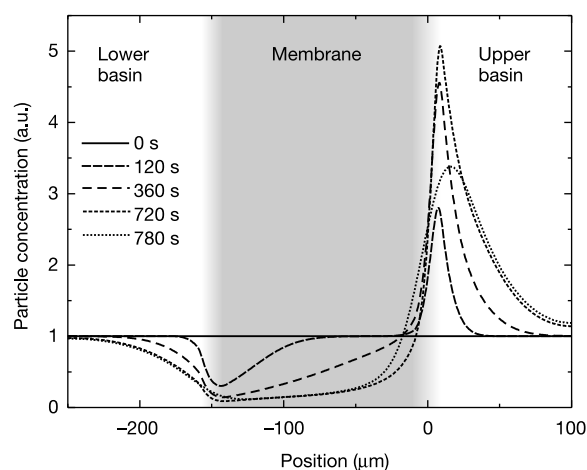


Figure 3 Simulated profile of particle concentration as a function of position for different instants in time. The particle concentration in arbitrary units was obtained by numerical integration of equation (2) using the same parameter values as in Fig. 2b. The membrane area is shaded in grey with smooth edges. As described in the text, the properties are gradually changed from the membrane to the basin to reflect the smoothing influence of the oscillations.

basin into the membrane are the reasons for the observed pseudo-saturation. The real saturation, where nearly all particles are in the upper basin, takes much longer because of the macroscopic dimensions of the lower basin and the slow particle diffusion. The occurrence of this pseudo-saturation in our measurement is a strong indication of the efficient transport inside the membrane. Only for drift values near the simulated $v = 0.5 \mu\text{m s}^{-1}$ do we expect the strong reduction in transport capacity to occur after only a few hundred seconds. In a real application device these problems could be overcome by introducing strong mixing in the basin.

The characteristic sawtooth-like profile does not in general occur for the lower basin (L) because the reduction of concentration in the lower basin itself is dominated by the diffusion process from the lower basin to the particle-depleted ratchets. The drift process removes the particles from the lower interface and, therefore, does not directly lead to a change in the number of particles in the lower basin. After a few cycles the concentration near the lower membrane surface is already reduced to a few per cent of the starting homogeneous concentration. It changes only slowly over time by the diffusion process of the particles in the lower basin and leads to the rather unstructured evolution observed in Fig. 2a (U_{reversed}).

To avoid long-term stability problems we investigated the pressure dependence with a more sophisticated pressure profile. Only a few 'on/off' cycles are performed for each amplitude, followed by a relaxation time of 200 s and then immediately followed by the next parameter set. To check reproducibility, we repeated a pressure amplitude leading to easily observable transport every fifth measurement (see Supplementary Information). The resulting dependence is depicted in Fig. 4. The error bars are estimated from the deviations in the linear fits. For low-pressure amplitudes the particles drift to the short end of the pore modulation (upper basin in Fig. 1, negative current). At 3,000 Pa the direction of the net motion switches its sign and the particles show a drift to the long end (lower basin in Fig. 1). By simply tuning the applied amplitude of the pressure oscillations the direction of transport can be reversed. For the positive direction, the experimental observation is expected to be much more difficult, as mentioned in the context of Fig. 2. Nevertheless, we can observe a few points with moderate error bars.

If the amplitude is close to zero, no transport will occur because the oscillating water-plus-particle system only moves a very short

distance compared to the length of one modulation and therefore does not feel the asymmetric pore shape. Also at high-pressure values the transport seems to diminish, as predicted by theory, although the error bars presented here do not really allow us to verify this. The experimentally observed pressure dependence strongly supports the underlying microscopic theory.

Our measurements are in qualitative agreement with the results of the theoretical calculations, suggesting that the predicted^{17,18} strong dependence of the direction of transport on particle size is likely to occur, and that the ratchet membrane might act as a massively parallel one-dimensional brownian separator¹⁸ for separation applications. Provided that problems associated with mixing and surface passivation can be resolved, the system seems well suited for the efficient and selective continuous separation of sensitive biological materials like viruses or cell fragments. As the external force is based on the flow resistance of the particles, it should be possible to separate elongated or flexible particles, provided their size is roughly in the 0.1–1 μm range. It is expected that the well developed electronic and micromechanical technology for silicon processing will allow the control of surface properties and the integration of many of the still discrete external components into 'lab on a chip' systems. In such highly integrated systems the dimensions could be drastically reduced, which would help to suppress the limiting diffusion processes in the basins and increase speed and sensitivity of the device considerably. $\square$

Methods

Sample fabrication

The electrochemical etching of silicon is carried out in HF ($c_{\text{HF}} = 5 \text{ wt\%}$; $T = 10^\circ\text{C}$; $U = 2 \text{ V}$) under backside illumination. The substrate is n-type FZ-Si (100) with a resistivity of $5 \Omega \text{ cm}^{-1}$. Using a photolithographic pre patterning, the macropores are arranged in a triangular lattice with a pitch of $a = 6 \mu\text{m}$. After electrochemical etching, a protective SiO_2 film is grown by thermal oxidation at 800°C in an oxygen atmosphere for 2 h. This leads to a SiO_2 thickness of 20 nm. After removing the SiO_2 -layer at the backside of the silicon wafer locally by an HF-dip, the backside is removed with an anisotropic KOH-etching process. The backside of the sample is exposed to 25 wt% KOH at 90°C . Before reaching the SiO_2 etch stop, temperature is reduced to 60°C . A well-defined hydrophilic OH-terminated surface is achieved via a standard cleaning process²³: the membrane is exposed to a mixture of $\text{H}_2\text{O}:\text{NH}_3:\text{H}_2\text{O}_2 = 5:1:1$ at 80°C for 10 min.

Particles

The functionalized particles (surfactant-free, fluorescent yellow-green sulphate polystyrene latex) are provided by IDC (Interfacial Dynamics Corporation). The supplied solution is diluted after ultrasonic treatment using ultra-pure deionized water from a Millipore Milli-RX20 ($18 \text{ M}\Omega \text{ cm}^{-1}$, filter pore diameter $0.22 \mu\text{m}$).

Measurements

Particle concentration is measured using PL. For the highly diluted solutions used, the PL intensity is proportional to the number of particles in the upper basin. Excitation is done with 488-nm light from an argon laser. The PL is detected at 515 nm using a Spex 270M monochromator equipped with a Hamamatsu photomultiplier tube R928. The pressure oscillations are achieved by a membrane pump SMF 4 from ASF Thomas Industries. The applied pressure is detected with a sensor from Honeywell 26PCCFRAID. The filling level in the upper basin is measured capacitively using a Boonton 7200.

Received 21 January 2002; accepted 13 May 2003; doi:10.1038/nature01736.

- Huxley, A. F. Muscle structure and theories of contraction. *Prog. Biophys. Biophys. Chem.* **7**, 255–318 (1957).
- von Smoluchowski, M. Experimentell nachweisbare, der üblichen Thermodynamik widersprechende Molekularphänomene. *Phys. Zeitsch.* **13**, 1069–1080 (1912).
- Feynman, R. P., Leighton, R. B. & Sands, M. *The Feynman Lectures On Physics* Ch. 46 Vol. 1 (Addison-Wesley, Reading, Massachusetts, 1966).
- Faucheux, L. P., Bourdieu, L. S., Kaplan, P. D. & Libchaber, A. J. Optical thermal ratchet. *Phys. Rev. Lett.* **74**, 1504–1507 (1995).
- Marquet, C., Buguin, A., Talini, L. & Silberzan, P. Rectified motion of colloids in asymmetrically structured channels. *Phys. Rev. Lett.* **88**, 168301 (2002).
- Derényi, I. & Astumian, R. D. AC separation of particles by biased Brownian motion in a two-dimensional sieve. *Phys. Rev. E* **58**, 7781–7784 (1998).
- Rousselet, J., Salome, L., Ajdari, A. & Prost, J. Directional motion of brownian particles induced by a periodic asymmetric potential. *Nature* **370**, 446–448 (1994).
- van Oudenaarden, A. & Boxer, S. G. Brownian ratchet: molecular separations in liquid bilayers supported on patterned arrays. *Science* **285**, 1046–1048 (1999).
- Linke, H. et al. Experimental tunneling ratchets. *Science* **286**, 2314–2317 (1999).
- Switkes, M., Marcus, C. M., Campman, K. & Gossard, A. C. An adiabatic quantum electron pump. *Science* **283**, 1905–1908 (1999).

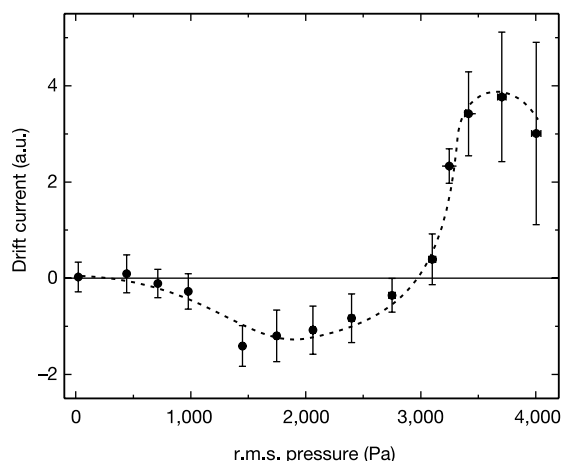


Figure 4 Drift current j_{drift} calculated from experimental results on particle transport as a function of applied pressure amplitudes. The transport of particles with a diameter of $0.1 \mu\text{m}$ diminishes for pressure values close to zero and large amplitudes. At an r.m.s. pressure of about 3,000 Pa the particle current is reversed. The direction of transport can be reversed by tuning the applied amplitude of the pressure.

11. Hänggi, P. & Bartussek, R. *Lecture Notes in Physics* Vol. 476 (eds Parisi, J., Müller, S. C. & Zimmermann, W.) 294–308 (Springer, Berlin, 1996).
12. Astumian, R. D. Thermodynamics and kinetics of a brownian motor. *Science* **276**, 917–922 (1997).
13. Astumian, R. D. & Hänggi, P. Brownian motors. *Phys. Today* **55**(11), 33–39 (2002).
14. Linke, H. (ed.) Ratchets and Brownian motors: basics, experiments and applications. *Appl. Phys. A* (Special Issue) **75**(2), (2002).
15. Cabodi, M., Chen, Y.-F., Turner, S. W. P., Craighead, H. G. & Austin, R. H. Continuous separation of biomolecules by the laterally asymmetric diffusion array with out-of-plane sample injection. *Electrophoresis* **23**, 3496–3503 (2002).
16. Müller, F. et al. Membranes for micropumps from macroporous silicon. *Phys. Stat. Sol. A* **182**, 585–590 (2000).
17. Kettner, C., Reimann, P., Hänggi, P. & Müller, F. Drift ratchet. *Phys. Rev. E* **61**, 312–323 (2000).
18. Kettner, C., Hänggi, P. & Müller, F. Verfahren und Vorrichtung zur größenabhängigen Sortierung mikroskopisch kleiner Teilchen auf der Basis von rauschinduziertem Transport. German patent pending (DE 199 07 564 A 1), submitted 22 Feb. 1999.
19. Lehmann, V. & Föll, H. Formation mechanism and properties of electrochemically etched trenches in n-type silicon. *J. Electrochem. Soc.* **137**, 653–659 (1990).
20. Lehmann, V. The physics of macropore formation in low doped n-type silicon. *J. Electrochem. Soc.* **140**, 2836–2843 (1993).
21. Schilling, J. et al. 3D photonic crystals made out of macroporous silicon by modulation of pore diameter. *Appl. Phys. Lett.* **78**, 1180–1182 (2001).
22. Lehmann, V. & Grüning, U. The limits of macropore array fabrication. *Thin Solid Films* **297**, 13–17 (1997).
23. Tong, Q. Y. & Gösele, U. *Semiconductor Wafer Bonding* 53–54 (Wiley and Sons, New York, 1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgement We thank U. Gösele for backing the ratchet project over many years. We also gratefully acknowledge discussions on the theory of the ratchet-effect with P. Hänggi, Ch. Kettner, P. Reimann and R. Eichhorn as well as critical reading of the manuscript by K. Scheersmidt and R. B. Wehrspohn.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.M. (Frank.Mueller@mpi-halle.de).

High $^3\text{He}/^4\text{He}$ ratios in picritic basalts from Baffin Island and the role of a mixed reservoir in mantle plumes

Finlay M. Stuart*, Solveigh Lass-Evans*†, J. Godfrey Fitton† & Robert M. Ellam*

* Isotope Geosciences Unit, Scottish Universities Environmental Research Centre, East Kilbride G75 0QF, UK

† School of GeoSciences, University of Edinburgh, Edinburgh EH9 3JW, UK

The high $^3\text{He}/^4\text{He}$ ratio of volcanic rocks thought to be derived from mantle plumes is taken as evidence for the existence of a mantle reservoir that has remained largely undegassed since the Earth's accretion^{1–3}. The helium isotope composition of this reservoir places constraints on the origin of volatiles within the Earth and on the evolution and structure of the Earth's mantle. Here we show that olivine phenocrysts in picritic basalts presumably derived from the proto-Iceland plume at Baffin Island, Canada, have the highest magmatic $^3\text{He}/^4\text{He}$ ratios yet recorded. A strong correlation between $^3\text{He}/^4\text{He}$ and $^{87}\text{Sr}/^{86}\text{Sr}$, $^{143}\text{Nd}/^{144}\text{Nd}$ and trace element ratios demonstrate that the ^3He -rich end-member is present in basalts that are derived from large-volume melts of depleted upper-mantle rocks. This reservoir is consistent with the recharging of depleted upper-mantle rocks by small volumes of primordial volatile-rich lower-mantle material at a thermal boundary layer between convectively isolated reservoirs. The highest $^3\text{He}/^4\text{He}$ basalts from Hawaii and Iceland plot on the observed mixing trend. This indicates that a ^3He -recharged depleted mantle (HRDM) reservoir may be the principal source

of high $^3\text{He}/^4\text{He}$ in mantle plumes, and may explain why the helium concentration of the 'plume' component in ocean island basalts is lower than that predicted for a two-layer, steady-state model of mantle structure.

Ocean island basalts (OIB) derived from mantle plumes commonly have $^3\text{He}/^4\text{He}$ ratios that are higher than mid-ocean ridge basalts (MORB) that originate in the upper, asthenospheric mantle^{1–3}. High $^3\text{He}/^4\text{He}$ and solar-like Ne isotope ratios⁴ reflect a higher proportion of primordial volatiles in the source region and, in the current paradigm, indicate that plumes tap a deep mantle reservoir that is significantly less degassed than the asthenospheric mantle. Although a number of locations have been proposed for this reservoir⁵, a lower mantle that has been convectively isolated below the 670-km seismic discontinuity for the lifetime of the Earth⁶ is consistent with the mass balance of the depleted mantle and continental crust⁷. However, layered mantle models are seemingly at odds with seismic studies that demonstrate convective flow across the 670-km boundary⁸ and require a 100- to 200-fold depletion of the He concentration in the upper, degassed mantle relative to the deep undegassed mantle reservoir⁶. Plume-derived OIB from, for example, Loihi seamount, Hawaii, have lower He concentrations than basalts from the degassed upper mantle² and have posed a persistent problem for a unified geochemical model of Earth structure⁶. The apparent paradox can be explained, at least in part, by more extensive degassing of ocean island basalts⁹. However, basalts from Iceland and the Hawaiian islands have linear, rather than strongly hyperbolic, He–Pb isotope mixing arrays that are consistent with relatively small He concentration contrasts between the degassed mantle and less degassed (high $^3\text{He}/^4\text{He}$) mantle reservoirs^{9,10}.

Picritic basalts erupted on Baffin Island, northeast Canada, 61 Myr ago correlate with the Anaanaa Member of the Vaigat Formation in West Greenland¹¹, and are among the earliest manifestations of the ancestral Iceland mantle plume¹². Trace-element and Sr–Nd isotope ratios show that they were generated by mixing of two discrete sources; relatively depleted mantle with $(\text{La}/\text{Sm})_n < 1$, $^{87}\text{Sr}/^{86}\text{Sr} \approx 0.7030$, $^{143}\text{Nd}/^{144}\text{Nd} \approx 0.5130$, and relatively enriched mantle with $(\text{La}/\text{Sm})_n > 1$, $^{87}\text{Sr}/^{86}\text{Sr} \approx 0.7042$ and $^{143}\text{Nd}/^{144}\text{Nd} \approx 0.51282$ (ref. 13). The depleted basalts are similar to the most depleted Icelandic basalts and North Atlantic MORB¹⁴. The enriched basalts are similar to, though slightly more extreme (higher $^{87}\text{Sr}/^{86}\text{Sr}$ and lower $^{143}\text{Nd}/^{144}\text{Nd}$ ratios) than the most enriched basalts erupted by the Iceland plume and across the North Atlantic igneous province¹⁵.

The isotopic composition of He released by *in vacuo* crushing olivine phenocrysts from picritic basalts from three stratigraphically distinct cliff sections in northeast Baffin Island (Table 1) ranges from $14.4R_a$ to $49.5R_a$ (where R_a is the atmospheric $^3\text{He}/^4\text{He}$ ratio; 1.39×10^{-6}). These ratios are far higher than MORB values and extend the range measured in contemporaneous picrites from West Greenland¹⁶. The high ratios cannot result from contamination by *in situ* cosmogenic ^3He because (1) *in vacuo* crushing does not release cosmogenic He from the olivine lattice¹⁷, (2) cosmogenic ^3He production was negligible because the samples were collected from eroding, near-vertical sea cliff faces, and (3) the $^3\text{He}/^4\text{He}$ of powders produced by crushing are systematically lower than crush values (Table 1) indicating that radiogenic He dominates the olivine lattice. Nucleogenic ^3He production from ^6Li is negligible in Palaeocene basalts¹⁶.

Crush release of radiogenic ^4He in the olivine lattice and melt inclusions by post-eruptive U and Th decay tends to lower $^3\text{He}/^4\text{He}$, so measured ratios may underestimate the source $^3\text{He}/^4\text{He}$. Any such effect would be most pronounced in samples with the lowest He content. $^3\text{He}/^4\text{He}$ are not correlated with He concentration (Table 1), implying that post-eruptive radiogenic He is insignificant to the crush-release He. Crustal contamination of the basalts by amphibolite- and granulites-facies basement rocks would also lower

Table 1 He, Sr and Nd isotope ratios, and trace-element data for Baffin Island basalts

Sample	$^4\text{He}^*$ ($10^{-9}\text{cm}^3\text{STP g}^{-1}$)	$^3\text{He}/^4\text{He}^*$ (R/R_a)	$^4\text{He}^\dagger$ ($10^{-9}\text{cm}^3\text{STP g}^{-1}$)	$^3\text{He}/^4\text{He}^\dagger$ (R/R_a)	$^{143}\text{Nd}/^{144}\text{Nd}\ddagger$	$^{87}\text{Sr}/^{86}\text{Sr}\ddagger$	(La/Sm) _n	ΔNb
Cape Searle								
CS/19	10.0 ± 0.3	23.0 ± 1.3			0.512924 ± 6	0.703420 ± 17	1.82	0.26
CS/7	130.6 ± 0.6	43.9 ± 0.6	33.6 ± 0.17	7.4 ± 0.5	0.512730 ± 11	0.704912 ± 17	0.92	-0.11
CS/6	20.9 ± 0.2	37.9 ± 1.2	6.9 ± 0.58	4.7 ± 0.7	0.512929 ± 7	0.703445 ± 20	1.78	0.20
CS/5	4.4 ± 0.2	26.3 ± 2.0	169.6 ± 0.86	0.16 ± 0.04	0.512921 ± 13	0.703314 ± 20	1.86	0.19
Padloping Island								
PI/23	39.8 ± 0.2	43.9 ± 0.6	23.4 ± 0.12	2.1 ± 0.1	0.513051 ± 8	0.703248 ± 20	1.00	-0.17
PI/25	11.0 ± 0.1	49.5 ± 1.6	164.7 ± 0.82	0.12 ± 0.02	0.513030 ± 6	0.703142 ± 24	1.04	-0.05
PI/26	24.1 ± 0.1	43.9 ± 0.5	6.6 ± 0.05	17.3 ± 0.8	0.513019 ± 6	0.703181 ± 24	1.15	-0.23
PI/27	45.6 ± 0.2	43.1 ± 0.5	11.8 ± 0.07	9.9 ± 1.2	0.513026 ± 8	0.703076 ± 21	1.00	-0.12
Durban Island								
DI/23	22.4 ± 0.3	14.5 ± 0.5	64.3 ± 0.33	2.1 ± 0.13	0.512845 ± 9	0.703964 ± 20	2.25	0.26
DI/24	5.1 ± 0.04	37.5 ± 0.6	6.2 ± 0.33	10.5 ± 0.8	0.513012 ± 6	0.703572 ± 20	1.22	-0.25
DI/27	17.4 ± 0.4	38.6 ± 1.5	7.0 ± 0.08	6.9 ± 0.8	0.512976 ± 8	0.703372 ± 21	0.93	-0.23

*Helium was extracted by *in vacuo* crushing 1–2 g of olivine and analysed at SUERC using procedures reported previously¹⁹. $^3\text{He}/^4\text{He}$ ratios (R) are normalized to the atmospheric ratio (1.39×10^{-6} ; R_a) and corrected for blanks. ^4He blanks averaged $4 \times 10^{-11}\text{cm}^3\text{STP}$ and ^3He blanks ($5 \times 10^{-15}\text{cm}^3\text{STP}$) were always less than 5% of the measured ^3He . Correction for atmospheric He on the basis of Ne abundances is insignificant.

†Olivine powders (~100 mg) produced by *in vacuo* crushing were melted at 1,850 °C in a double-walled resistance furnace. Hot furnace ^4He blanks averaged $2 \times 10^{-11}\text{cm}^3\text{STP g}^{-1}$. ^3He blanks were the same as crush extractions.

‡Whole-rock Sr and Nd isotopic ratios²⁸ have been corrected to 60 Myr ago on the basis of measured Rb/Sr and Sm/Nd. This reduces $^{87}\text{Sr}/^{86}\text{Sr}$ by less than 0.007% and $^{143}\text{Nd}/^{144}\text{Nd}$ by less than 0.002%. Uncertainties (2σ) are in the last decimal place. Rare-earth element concentrations were measured by ICP-MS²⁹. (La/Sm)_n is chondrite normalized, and has an uncertainty of ± 3% (2σ). Nb, Y and Zr concentrations were measured in triplicate by X-ray fluorescence and the averages used²⁴. ΔNb = $1.74 + \log(\text{Nb}/\text{Y}) - 1.92 \times \log(\text{Zr}/\text{Y})$ ³⁰. Uncertainty in ΔNb never exceeds ± 10%.

$^3\text{He}/^4\text{He}$. Pb isotopes are sensitive to crustal contamination in North Atlantic margin Palaeocene basalts, because the Precambrian basement has distinctively unradiogenic Pb (ref. 18). The Baffin Island picrites show no correlation between Pb isotope composition and other isotopic tracers (Sr, Nd, He). Thus, while the Pb isotopes may indicate some crustal interaction, its effect on other elements appears negligible. We have previously demonstrated that basalts whose whole-rock Pb isotope compositions are clearly affected by crustal contamination retain uncontaminated He (ref. 19), either because the olivine crystallized before crustal assimilation, or because the deep crust is He-poor. In this study, we observe a similar effect: dyke sample CS/7 has high $^3\text{He}/^4\text{He}$ ($43.3 \pm 0.6R_a$) but has high $^{87}\text{Sr}/^{86}\text{Sr}$ and low $^{143}\text{Nd}/^{144}\text{Nd}$ that are best attributed to crustal contamination. As crustal contamination is expected to lower $^3\text{He}/^4\text{He}$, we conclude that crush-release He is insensitive to any contamination. The inability of magmatic or post-magmatic processes to alter He isotope ratios, and the strong correlation with Sr and Nd isotopes and trace elements (see below), means that the sample (PI/25) with the highest helium isotope ratio provides the best estimate yet ($49.5 \pm 1.5R_a$) of the minimum He isotope composition of the less degassed mantle reservoir.

The olivine $^3\text{He}/^4\text{He}$ ratio correlates strongly with whole-rock $^{143}\text{Nd}/^{144}\text{Nd}$, $^{87}\text{Sr}/^{86}\text{Sr}$ and trace-element ratios such as La/Sm (see Fig. 1). The absence of co-variation of $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ with indices of contamination such as Ba/Nb and Pb isotope ratios suggests that crustal contamination has not significantly affected Sr and Nd isotopes (with the exception of the dyke sample CS/7). Furthermore, high- $^3\text{He}/^4\text{He}$ Palaeocene basalts from Iceland³ that cannot have been contaminated by continental crust have He–Nd and He–Sr isotopic compositions that plot on the trend. Consequently, the He isotope correlations with La/Sm and the lithophile isotope ratios record binary mixing between large-volume melts of a ^3He -rich, relatively depleted mantle reservoir, and small-volume melts from a mantle reservoir with low $^3\text{He}/^4\text{He}$, rich in incompatible elements.

The depleted mantle end-member inferred from the most extreme Sr and Nd isotope compositions requires a mantle source region with $^{87}\text{Sr}/^{86}\text{Sr} \approx 0.70305$, $^{143}\text{Nd}/^{144}\text{Nd} \approx 0.51305$ and $^3\text{He}/^4\text{He} > 50R_a$. The Sr and Nd isotope ratios are indistinguishable from the most depleted basalts erupted on Iceland and only slightly less depleted than North Atlantic MORB erupted away from the Iceland plume ($^{87}\text{Sr}/^{86}\text{Sr} < 0.7030$, $^{143}\text{Nd}/^{144}\text{Nd} > 0.5131$). Icelandic and North Atlantic MORBS define parallel, non-overlapping arrays on a plot of $\log(\text{Nb}/\text{Y})$ against $\log(\text{Zr}/\text{Y})$. The excess or

deficiency in Nb is defined by the ΔNb notation such that Icelandic basalts have ΔNb > 0 and MORB have ΔNb < 0 (ref. 24). The high $^3\text{He}/^4\text{He}$ Baffin Island basalts have strongly negative ΔNb (Table 1) indicating derivation from the North Atlantic MORB-source mantle rather than the depleted Iceland plume.

It is, first, important to note that this He-, Sr- and Nd-isotopic composition cannot represent MORB source mantle at 61 Myr ago as the $^3\text{He}/(\text{U}+\text{Th})$ of the MORB reservoir is incapable of significantly changing $^3\text{He}/^4\text{He}$ in hundreds of millions of years. If He is less incompatible than U during melting, high- $^3\text{He}/^4\text{He}$ would correspond with the most depleted basalts²⁰. Low- $^{87}\text{Sr}/^{86}\text{Sr}$, high- $^{143}\text{Nd}/^{144}\text{Nd}$ and high- $^3\text{He}/^4\text{He}$ basalts are indicative of a

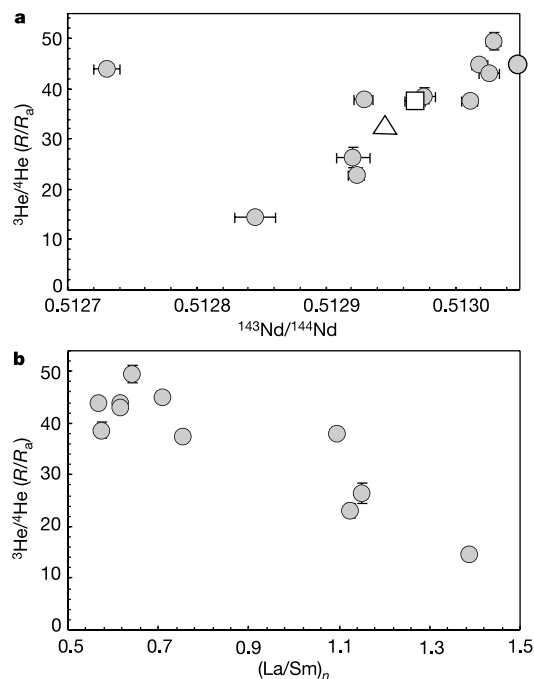


Figure 1 Plots of olivine $^3\text{He}/^4\text{He}$ versus $^{143}\text{Nd}/^{144}\text{Nd}$ and (La/Sm)_n for early Palaeocene basalts from Baffin Island. **a**, Whole-rock Nd isotopic composition; also shown are the data for a Palaeocene ankaramite from NW Iceland (square) and basaltic glass from Loihi seamount (triangle). **b**, Whole-rock (La/Sm)_n.

depleted, 'regassed' component which would be generated in a reservoir with low U/He which grows ^4He less quickly than MORB source mantle. The mechanism proposed to generate the low-U/He source²⁰ requires (shallow) melt-gas separation to fractionate He from U and the other lithophile trace elements. It is difficult to reconcile the linear He–Sr and He–Nd isotope arrays (Fig. 1) with such a component since a high-He/Sr and -He/Nd ratio component mixed with 'normal' He/Sr and He/Nd mantle would result in strongly hyperbolic mixing arrays. Instead, we make the assumption that the dominant control on isotopic heterogeneity in the mantle is the same for all elements—that is, solid-melt distribution coefficients—and that He is highly incompatible during mantle melting²¹. Therefore any reservoir with low Rb/Sr and high Sm/Nd, because it is residual to partial melting, will have lost most of its He and would evolve low- $^3\text{He}/^4\text{He}$, low- $^{87}\text{Sr}/^{86}\text{Sr}$ and high- $^{143}\text{Nd}/^{144}\text{Nd}$ ratios. The subduction of interplanetary dust particles is incapable of delivering enough solar noble gases to the upper mantle to change its He and Ne isotopic composition significantly²². In the absence of alternatives we show below that the high- $^3\text{He}/^4\text{He}$ component is most probably the product of mantle mixing.

If Earth accreted with $^3\text{He}/^4\text{He} = 120R_a$, bulk silicate Earth U (21 p.p.b.) and U/Th = 3.8, an undegassed mantle reservoir with $^3\text{He}/^4\text{He} = 50R_a$ has a ^3He concentration of 8×10^{10} atoms g^{-1} . If moderately incompatible lithophile element concentrations in the depleted mantle reservoir are less than a factor of two lower than undepleted mantle²³, mixing a small proportion of ^3He -rich mantle with depleted upper mantle ($1.2\text{--}4.6 \times 10^9$ ^3He atoms g^{-1} ; ref. 6) severely changes $^3\text{He}/^4\text{He}$ without significantly altering trace element concentrations and Sr and Nd isotope ratios. Consequently the isotopic composition of the high $^3\text{He}/^4\text{He}$ end-member can be generated by mixing depleted mantle and primordial ^3He -rich mantle reservoir. The Baffin Island data define approximately linear mixing arrays in He–Sr and He–Nd isotope space, implying that the He/Nd and He/Sr ratios and, therefore, He concentrations, of both reservoirs are similar. The absence of strongly hyperbolic mixing trends supports the contention that the high $^3\text{He}/^4\text{He}$ end-member is not simply undegassed mantle.

The enriched mantle end-member, estimated by extrapolating the He–Nd and He–Sr isotope trends to the most extreme values measured in uncontaminated Baffin Island basalts ($^{143}\text{Nd}/^{144}\text{Nd} = 0.51282$; $^{87}\text{Sr}/^{86}\text{Sr} = 0.7042$), has a $^3\text{He}/^4\text{He}$ similar to, or lower than, the MORB ratio ($8 \pm 1 R_a$). The positive Nb value of this end-member, and the coincidence of the high $^3\text{He}/^4\text{He}$ ankaramite from northwest Iceland³ with the mixing trend, supports an origin for this enrichment in the mantle asthenosphere. Although the ultimate origin is unclear, the low $^3\text{He}/^4\text{He}$ ratio requires that it has been isolated from the high- $^3\text{He}/^4\text{He}$ reservoir prior to melting. It may represent fusible heterogeneities in the upper mantle or the plume that are large enough to prevent isotopic homogenization yet small enough to be preferentially sampled by melting.

The convergence of Sr–Nd–Pb isotope trends in ocean island basalt (OIB) and MORB have been used to identify a common deep mantle reservoir that is the carrier of the high- $^3\text{He}/^4\text{He}$ plume component. The most primordial ^3He -rich basalts from Iceland and Hawaii lie on the He–Nd and He–Sr isotope mixing trends identified here suggesting that the extreme end of the Baffin Island picrite trend is the best estimate of a common, high- $^3\text{He}/^4\text{He}$ mantle component. This is at the depleted end of the ranges proposed previously for a common mantle reservoir^{24,25}, but the negative ΔNb requires that it originates in the depleted upper mantle, rather than the lower mantle. The thermal boundary layer between upper- and lower-mantle regions is one potential reservoir for such material^{26,27}, although it cannot be ruled out that the mixing occurs as the plume head disperses in the upper mantle. The dominance of the ^3He -recharged, depleted component in mantle plumes is consistent with a low mass flux from the lower mantle and explains why

basalts from Iceland and Hawaii appear to have lower He concentrations than predicted for undegassed mantle^{9,10}. □

Received 12 December 2002; accepted 29 April 2003; doi:10.1038/nature01711.

- Kurz, M. D., Jenkins, W. J. & Hart, S. R. Helium isotope systematics of ocean islands and mantle heterogeneity. *Nature* **297**, 43–46 (1982).
- Honda, M., McDougall, I., Patterson, D. B., Douglis, A. & Clague, D. A. Noble gases in submarine pillow basalt glasses from Loihi and Kilauea, Hawaii—a solar component in the Earth. *Geochim. Cosmochim. Acta* **57**, 859–874 (1993).
- Hilton, D. R., Gronvold, K., Macpherson, C. & Castillo, P. Extreme $^3\text{He}/^4\text{He}$ ratios in northwest Iceland: constraining the common component in mantle plumes. *Earth Planet. Sci. Lett.* **173**, 53–60 (1999).
- Dixon, E. T., Honda, M., McDougall, I., Campbell, I. H. & Sigurdsson, I. Preservation of near-solar neon isotopic ratios in Icelandic basalts. *Earth Planet. Sci. Lett.* **180**, 309–324 (2000).
- Porcelli, D. & Ballentine, C. J. Origin of noble gases in terrestrial planets. *Rev. Mineral. Geochem.* **47**, 411–480 (2002).
- Porcelli, D. & Wasserburg, G. J. Mass transfer of helium, neon, argon, and xenon through a steady-state upper mantle. *Geochim. Cosmochim. Acta* **59**, 4921–4937 (1995).
- DePaolo, D. J. & Wasserburg, G. J. Inferences about magma sources and mantle structure from variations in $^{143}\text{Nd}/^{144}\text{Nd}$. *Geophys. Res. Lett.* **3**, 743–746 (1976).
- van der Hilst, R. D., Widiyandanto, S. & Engdahl, E. R. Evidence for deep mantle circulation from global tomography. *Nature* **386**, 578–584 (1997).
- Hilton, D. R., Thirlwall, M. F., Taylor, R. N., Murton, B. J. & Nichols, A. Controls on magmatic degassing along the Reykjanes Ridge with implications for the helium paradox. *Earth Planet. Sci. Lett.* **183**, 43–50 (2000).
- Eiler, J. M., Farley, K. A. & Stolper, E. M. Correlated He and Pb isotope variations in Hawaiian lavas. *Geochim. Cosmochim. Acta* **63**, 1977–1998 (1998).
- Pedersen, A. K., Larsen, L. M., Riisager, P. & Dueholm, K. S. In *The North Atlantic Igneous Province: Stratigraphy, Tectonic, Volcanic and Magmatic Processes* (eds Jolley, D. W. & Bell, B. R.) 157–181 (Geol. Soc. Lond. Spec. Publ. 197, 2002).
- Francis, D. The Baffin Bay lavas and the value of picrites as analogues of primary magmas. *Contrib. Mineral. Petrol.* **89**, 144–154 (1985).
- Robillard, L., Francis, D. & Ludden, J. N. The relationship between E- and N-type magmas in the Baffin Bay lavas. *Contrib. Mineral. Petrol.* **112**, 230–241 (1989).
- Saunders, A. D., Fitton, J. G., Kerr, A. C., Norry, M. J. & Kent, R. W. In *Large Igneous Provinces* (eds Coffin, M. F. & Mahoney, J. J.) 45–94 (Am. Geophys. Union, Washington DC, 1997).
- Prestvik, T., Goldberg, S., Karlsson, H. & Gronvold, K. Anomalous strontium and lead isotope signatures in the off-rift Oraefajokull central volcano in south-east Iceland: Evidence for enriched end-member(s) of the Iceland plume? *Earth Planet. Sci. Lett.* **190**, 211–220 (2001).
- Graham, D. W. *et al.* Helium isotope composition of the early Iceland mantle plume inferred from the Tertiary picrites of West Greenland. *Earth Planet. Sci. Lett.* **160**, 241–255 (1998).
- Kurz, M. D. Cosmogenic helium in a terrestrial igneous rock. *Nature* **320**, 435–439 (1986).
- Whitehouse, M. J., Kalsbeek, F. & Nutman, A. P. Crustal growth and crustal recycling in the Nugsugtoquidian orogen of West Greenland: constraints from radiogenic isotope systematics and U–Pb zircon geochronology. *Precambrian Res.* **91**, 365–381 (1998).
- Stuart, F. M., Ellam, R. M., Harrop, P. J., Fitton, J. G. & Bell, B. R. Constraints on mantle plumes from the helium isotope composition of basalts from the British Tertiary Igneous Province. *Earth Planet. Sci. Lett.* **177**, 273–285 (2000).
- Anderson, D. A model to explain the various paradoxes associated with mantle noble gas geochemistry. *Proc. Natl Acad. Sci. USA* **95**, 9087–9092 (1998).
- Marty, B. & Lussiez, P. Constraints on rare gas partition coefficients from analysis of olivine-glass from a picritic mid-ocean ridge basalt. *Chem. Geol.* **106**, 1–7 (1993).
- Stuart, F. M. Comment on "Speculations about the cosmic origin of He and Ne in the interior of the Earth". *Earth Planet. Sci. Lett.* **122**, 245–247 (1994).
- McKenzie, D. M. & O'Nions, R. K. The source regions of ocean island basalts. *J. Petrol.* **36**, 133–159 (1995).
- Hanan, B. B. & Graham, D. W. Lead and helium isotope evidence from oceanic basalts for a common deep source of mantle plumes. *Science* **272**, 991–995 (1996).
- Hauri, E. H., Whitehead, J. A. & Hart, S. R. Fluid dynamic and geochemical aspects of entrainment in mantle plumes. *J. Geophys. Res.* **99**, 24275–24300 (1994).
- Fitton, J. G., Saunders, A. D., Norry, M. J., Hardarson, B. S. & Taylor, R. N. Thermal and chemical structure of the Iceland plume. *Earth Planet. Sci. Lett.* **153**, 197–208 (1997).
- Marty, B., Pik, R. & Gerzhanov, Y. Helium isotopic variation in Ethiopian plume lavas: nature of magma sources and limit on lower mantle contribution. *Earth Planet. Sci. Lett.* **144**, 223–237 (1996).
- Hardarson, B. S., Fitton, J. G., Ellam, R. M. & Pringle, M. Rift relocation—a geochemical and geochronological investigation of a palaeo-rift in northwest Iceland. *Earth Planet. Sci. Lett.* **153**, 181–196 (1997).
- Olive, V., Ellam, R. M. & Wilson, L. A protocol for the determination of the rare Earth elements at picomole level in rocks by ICP-MS: Results on geological reference materials USGSPCC-1 and DTS-1. *Geostand. Newsl. J. Geostand. Geanal.* **25**, 219–228 (2001).
- Fitton, J. G., Saunders, A. D., Larsen, L. M., Hardarson, B. S. & Norry, M. J. Volcanic rocks from the southeast Greenland margin at 63°N: composition, petrogenesis and mantle sources. *Proc. ODP Sci. Res.* **152**, 331–350 (1998).

Acknowledgements The samples used in this study were collected by I. Snape, C. Cole and S. Brown during an expedition partly funded by the Sheila Hall Trust. This work was supported by NERC scholarship to S.L.-E. and NERC funding to F.M.S. and R.M.E.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.M.S. (f.stuart@suerc.gla.ac.uk).

Carbon loss by deciduous trees in a CO₂-rich ancient polar environment

Dana L. Royer⁺, Colin P. Osborne & David J. Beerling

Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

Fossils demonstrate that deciduous forests covered the polar regions for much of the past 250 million years¹ when the climate was warm and atmospheric CO₂ high². But the evolutionary significance of their deciduous character has remained a matter of conjecture for almost a century³. The leading hypothesis^{1,4–7} argues that it was an adaptation to photoperiod, allowing the avoidance of carbon losses by respiration from a canopy of leaves unable to photosynthesize in the darkness of warm polar winters^{8–11}. Here we test this proposal with experiments using ‘living fossil’ tree species grown in a simulated polar climate with and without CO₂ enrichment. We show that the quantity of carbon lost annually by shedding a deciduous canopy is significantly greater than that lost by evergreen trees through wintertime respiration and leaf litter production, irrespective of growth CO₂ concentration. Scaling up our experimental observations indicates that the greater expense of being deciduous persists in mature forests, even up to latitudes of 83°N, where the duration of the polar winter exceeds five months. We therefore reject the carbon-loss hypothesis as an explanation for the deciduous nature of polar forests.

Plant fossils provide tangible evidence for the widespread occurrence of deciduous forests in the polar regions throughout the Mesozoic and Palaeogene¹. These high-latitude forests flourished in a CO₂-rich atmosphere², under winter temperatures considerably warmer than those of today^{8–11} and more closely resembling the northern coast of the Mediterranean¹² (Fig. 1). The carbon loss hypothesis^{1,4–7} maintains that the quantity of carbon lost when a canopy of leaves is shed annually (that is, the deciduous habit) is less than the carbon losses by canopy respiration during the warm, dark winter months and the annual abscission of only a fraction of leaves (that is, the evergreen habit). However, because no modern analogues exist for these ancient polar environments, the hypothesis remains untested. Experiments offer the most direct means of testing, but are necessarily shorter in duration than the lifetime of the trees. To address differences in carbon budgets over longer timescales applicable to mature forests, and across a wide latitudinal gradient, requires application of suitable process-based models of forest biogeochemistry. Here, therefore, we present an integrated study that tests the carbon-loss hypothesis by combining plant-growth experiments in a simulated high-latitude environment¹³ with numerical modelling simulations of conifer forests¹⁴.

We ran our experiments using three deciduous species (the taxodiaceous conifers *Metasequoia glyptostroboides* Hu & Cheng and *Taxodium distichum* (L.) Rich, and the broadleaved gymnosperm *Ginkgo biloba* L.) and two evergreen species (the taxodiaceous *Sequoia sempervirens* (D. Don) Endl., and the southern-hemisphere angiosperm *Nothofagus cunninghamii* (Hook.) Oerst.). All of these represent reasonable modern analogues for their congeneric ancestors, which were dominant elements in Cretaceous and early Palaeogene polar landscapes up to a latitude of 85°N (refs 1, 6). One-year-old saplings of each species were grown in a high-latitude (69°N) photoperiod and a simulated Cretaceous/early Palaeogene seasonal temperature regime for three years, in atmospheres of either 400 p.p.m.v. or 800 p.p.m.v.

CO₂ (Fig. 1, see ref. 13 for details). Because plant respiration typically increases with temperature¹⁵, a severe thermal challenge to the carbon-loss hypothesis was realized by experimentally imposing the warmest of the cold-month mean temperatures reconstructed from fossil biota^{8–11} (Fig. 1). Whole-tree net carbon exchange rates were tracked throughout the year using flow-through chambers, and the production and carbon content of leaf litter measured in the autumn (see Methods).

During the six weeks of continuous darkness in both the 2001–2002 and 2002–2003 simulated polar winters, carbon losses by leaf and stem respiration in the evergreen species were consistently small (Fig. 2a), equivalent to only 1–2% of annual net primary productivity (NPP) (Fig. 2d). Nevertheless, they were significantly larger than those of the leafless stems of the deciduous species (2002–2003 winter: $F_{1,6} = 90$, $P < 0.001$; 2001–2002 winter: $F_{1,6} = 140$, $P < 0.001$). Importantly, all of the trees survived three successive polar winters, in agreement with an earlier study¹⁶, and maintained a normal rhythm of leaf expansion and abscission in each growing season. The quantity of carbon lost in the production of deciduous leaf litter was nearly an order of magnitude greater than the comparatively minor litter turnover in evergreens ($F_{1,6} = 45$, $P < 0.001$) (Fig. 2b). No significant effects of CO₂ concentration in the growth chambers on either respiration or litter production were evident.

An initial evaluation of the carbon-loss hypothesis can be achieved by summing the quantity of carbon lost in leaf litter production and wintertime respiration for trees with each leaf habit (Fig. 2a,b). These figures demonstrate that all three deciduous tree species lost almost an order of magnitude more carbon than their evergreen counterparts in the same environmental setting (Fig. 2c), despite temperatures above freezing throughout the winter (Fig. 1). For the deciduous species, the loss of carbon originally invested into leaf production (Fig. 2b), along with the wintertime respiration of leafless stems (Fig. 2a), was a substantial fraction of the NPP (Fig. 2d) (14–25%), and significantly more than in the evergreens (1–3%) ($F_{1,6} = 59$, $P < 0.001$). This contrast is well illustrated by the closely related evergreen *S. sempervirens* and deciduous *M. glyptostroboides*, which both attained similar NPPs (Fig. 2d). The *M. glyptostroboides* leaf canopy lost 16 times more carbon than that of *S. sempervirens* at a CO₂ concentration of 400 p.p.m.v., and 14 times more at 800 p.p.m.v. (Fig. 2c). Respiration accounted for 90% of this loss in *S. sempervirens*, whereas litter production was 98% of the loss in *M. glyptostroboides*.

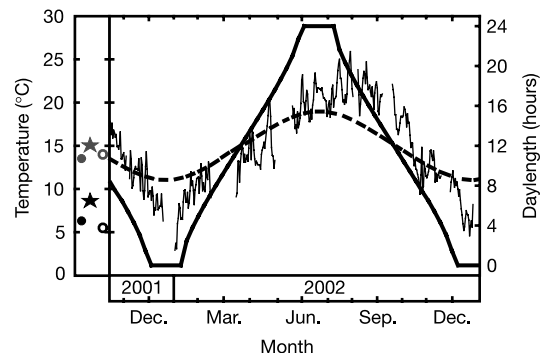


Figure 1 Environmental data for growth rooms. The thin line shows daily mean temperatures 2001–2002, and symbols in the left-hand column correspond to proxy estimates of coldest month mean temperature (black symbols) and mean annual temperature (grey symbols). Closed circles are estimates from geochemical analyses of early Palaeogene high-latitude marine molluscs^{10,11}, open circles are estimates from Cretaceous and early Palaeogene vertebrates^{8,9}, and stars are corresponding mean temperatures inside the growth rooms. The thick line denotes the simulated photoperiod at 69°N, and the dashed line shows for comparison that at 45°N.

⁺ Present address: Department of Geosciences and Institutes of the Environment, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

An evaluation of the relative merits of leaf habit on the carbon budgets of mature forests can be obtained by scaling up our measurements in space and time. The wintertime respiratory burden of an evergreen canopy, avoided by deciduous trees, will scale with canopy size, as described by the leaf area index (LAI, m^2 leaf area per m^2 ground area). In contrast, evergreens have the advantage of a substantially reduced rate of leaf litter production, to an extent dependent on the number of leaves produced each year, foliage longevity and tree age. We assessed the importance of our experimental findings for mature forests by addressing these two scaling issues and calculating the carbon cost (CC, gC per m^2 ground area) incurred by evergreen forests in a polar environment as $\text{CC} = \text{LAI} \times (R + \text{LCD} \times \tau)$, where R is the canopy respiration rate integrated for the period of continuous darkness (gC per m^2 leaf area), LCD is litter carbon density (gC per m^2 leaf area) and τ is the fraction of the canopy shed annually (dimensionless) (see Methods). For deciduous trees, where $R = 0$ and $\tau = 1$, the calculation of CC simplifies to $\text{CC} = \text{LAI} \times \text{LCD}$. Our analyses assume that the energetic costs of leaf construction for evergreen and deciduous species are similar, in agreement with observations¹⁷.

According to these scaling procedures, the difference in carbon cost between evergreen and deciduous forest stands is diminished relative to that measured directly in our experiments (Fig. 3a,b) because the scaled calculations more completely account for carbon losses by leaf fall from an evergreen canopy. Nevertheless, the cost of producing a deciduous canopy of leaves remains more than twice that incurred by evergreen trees through canopy respiration and turnover (Fig. 3a). Growth in a CO_2 -enriched atmosphere enhances the difference in carbon costs between leaf strategies (Fig. 3b) through a significant increase in LCD ($F_{1,30} = 9.6$, $P = 0.004$). Since there was no significant effect of CO_2 treatment on leaf tissue carbon concentration (gC per g leaf tissue) ($F_{1,30} = 0.51$, $P = 0.48$), the rise in LCD was driven solely by a significant increase in leaf mass per unit area (g dry matter per m^2 leaf tissue) ($F_{1,30} = 9.7$, $P = 0.004$), in agreement with other CO_2 studies on forest species at lower latitudes¹⁸. If this effect operated in ancient polar forests at times of increased atmospheric CO_2 , it may have exacerbated the carbon cost of having short-lived leaves.

Our experiments and scaling calculations are directly applicable

to a latitude of 69°N . However, the carbon-loss hypothesis predicts a greater penalty for evergreen trees at higher latitudes as cumulative canopy respiration increases with the duration of the polar winter. To investigate this issue at the scale of mature forests, we performed a series of simulations using a process-based numerical model of forest biogeochemistry¹⁴ forced with a Cretaceous climate and geography¹⁹ encompassing a range of latitudes (68 – 83°N ; see Methods) where mean temperatures were above freezing during winter darkness (Fig. 3c). In agreement with expectation, we find that wintertime respiration increases with latitude, despite the cooler temperatures, because of the lengthening polar winter (Fig. 3c,d). Nevertheless, this adverse effect on evergreens is negligible relative to the cost of annual canopy turnover in mature trees. As a consequence, the quantity of carbon required by deciduous conifers for annual canopy replacement continues to remain 50–100% greater, regardless of latitude (Fig. 3).

Proxy CO_2 data and geochemical modelling indicate a CO_2 -rich atmosphere for much of the Jurassic and Cretaceous periods², an interval when polar forests flourished and produced wide maximum growth ring widths²⁰. In our experiment, elevated CO_2 significantly ($F_{1,20} = 6.6$, $P = 0.02$) stimulated whole-tree NPP by between 17 and 73% (Fig. 2d). High CO_2 therefore might have partially compensated for the short high-latitude growing season experienced by these forests. Comparison of the NPP values for evergreen and deciduous taxa indicates that, despite a higher carbon cost, deciduous trees have a productivity similar to that of evergreens. This suggests that deciduous trees compensate for higher canopy carbon losses through increased rates of photosynthetic carbon gain.

Our results overturn the long-standing and pervasive conjecture that leaf habit in ancient polar forests was an adaptation to photoperiod. Moreover, it has been recognized that Antarctic polar forests appear to have contained a higher proportion of evergreens than their Arctic counterparts^{7,21,22}, a disparity also incompatible with the carbon-loss argument. We suggest that alternative explanations for the dominance of deciduous trees in these ecosystems must now be sought^{23,24}. The categorization of fossil forests as either evergreen or deciduous appears to represent an overly simplistic dichotomy in this context, since it assumes

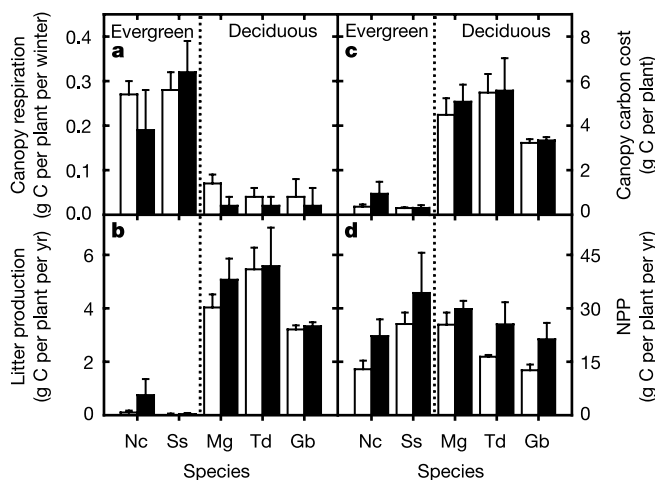


Figure 2 Experimental analyses of the carbon costs of leaf habit at 69°N for two atmospheric CO_2 levels. Open bars, 400 p.p.m.v. ; closed bars, 800 p.p.m.v. **a**, Canopy respiration during the 6-week period of continuous darkness of the 2002–2003 winter. **b**, Litter production in 2002. **c**, Total carbon costs of the polar winter, defined as the sum of **a** and **b**. **d**, Annual net primary productivity (NPP). Values are means of four replicates $\pm 1 \text{ s.e.}$ ($n = 3$ for NPP) for these species: Nc, *Nothofagus cunninghamii*; Ss, *Sequoia sempervirens*; Mg, *Metasequoia glyptostroboides*; Td, *Taxodium distichum* and Gb, *Ginkgo biloba*.

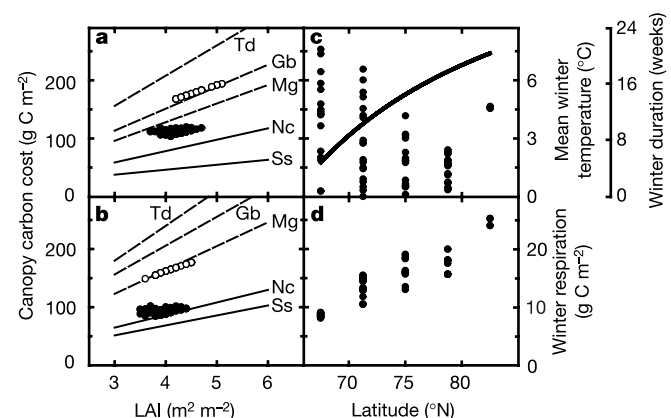


Figure 3 High-latitude forest carbon budgets and climates. Carbon costs of a polar winter for mature evergreen and deciduous leaf canopies calculated by scaling experimental observations (dashed lines, deciduous species; solid lines, evergreen species, defined as in Fig. 2) with leaf area index (LAI) for two atmospheric CO_2 levels. **a**, 400 p.p.m.v. and **b**, 800 p.p.m.v. Also shown are stand-level forest model simulations using a mid-Cretaceous climate. Open circles, leaf lifespan = 6 months; closed circles, leaf lifespan = 60 months across a range of latitudes (68 – 83°N). **c**, Mean Cretaceous climate model temperatures of the winter dark period (data points) and length of polar winter (solid line). **d**, Modelled wintertime canopy respiration of evergreen trees (leaf lifespan = 60 months).

similar function in all evergreen leaves, irrespective of their longevity. In fact, studies across biomes with different evolutionary and climatic histories demonstrate that this categorization is unlikely to be correct because leaf lifespan is strongly correlated with physiological, whole-tree and ecosystem processes^{14,25,26}. A palaeontological view of leaf lifespan is now afforded through anatomical analyses of growth rings in fossil woods²⁷, providing a critical means of linking past and present forest physiology. □

Methods

Whole-tree CO₂ flux measurements

The net carbon exchange rates of trees were tracked throughout the year under growth conditions using eight custom-built chambers, based on a modified design of ref. 28. Three replicate plants per CO₂ treatment were used, and the remaining two chambers used for controls, one from each CO₂ treatment, consisting of a pot with the same soil mixture as those with plants. Hourly measurements were made for each species over the course of a 24-h period, and repeated ten times throughout the year. Daily carbon budgets were estimated from the resulting 400 measurements, and then integrated to produce annual budgets (NPP) after taking into account the annual carbon loss by litter production. The carbon contribution of litter was estimated in sub-samples of leaf litter, using a stable-isotope ratio mass spectrometer (PDZ Europa 20–20, Cheshire).

CO₂ effects were analysed using a two-way ANOVA with replication (CO₂ × species); leaf habit effects were analysed using a two-way ANOVA with unequal but proportional subclass sizes (CO₂ × leaf habit) (ref. 29). None had significant interaction terms.

Scaling experimental results to mature stands

The following values for τ and measured LCD (mean ± s.e.) were used: for *N. cunninghamii* (at 400 p.p.m.v. CO₂: $\tau = 0.33$, LCD = 51.2 ± 2.6; at 800 p.p.m.v. CO₂: 0.33, 56.2 ± 7.1), for *S. sempervirens* (at 400 p.p.m.v. CO₂: 0.2, 44.0 ± 3.8; at 800 p.p.m.v. CO₂: 0.2, 70.1 ± 12.5), for *M. glyptostroboides* (at 400 p.p.m.v. CO₂: 1.0, 34.1 ± 2.6; at 800 p.p.m.v. CO₂: 1.0, 42.3 ± 3.0), for *T. distichum* (at 400 p.p.m.v. CO₂: 1.0, 53.0 ± 3.3; at 800 p.p.m.v. CO₂: 1.0, 60.4 ± 10.3) and for *G. biloba* (at 400 p.p.m.v. CO₂: 1.0, 38.9 ± 2.8; 800 p.p.m.v. CO₂: 1.0, 52.0 ± 1.7). See Fig. 2a for canopy respiration data (R). Measured rates of belowground respiration did not differ significantly between the two leaf habits during the polar winter ($F_{1,6} = 0.32$, $P = 0.59$), and so were not included in the scaling procedures.

Conifer forest modelling

The University of Sheffield Conifer Model (USCM) uses generalized relationships between leaf lifespan and function, and incorporates a full set of responses to atmospheric CO₂ and climate, as well as feedbacks with soil water and nutrient content¹⁴. USCM predictions of key properties and processes of conifer forests (NPP, LAI, evapotranspiration, nitrogen uptake and carbon partitioning) are validated well for forest sites across a wide climatic gradient using monthly climate data (temperature, precipitation and relative humidity), soil nutrient status and leaf lifespan information as inputs¹⁴. We forced the USCM with monthly climate data from the mid-Cretaceous climate simulation by the UK Universities Global Atmospheric Modelling Programme general circulation model¹⁹. Soil nutrient data for the Cretaceous period were derived using the Century soil biogeochemistry routines³⁰.

Received 28 February; accepted 22 April 2003; doi:10.1038/nature01737.

- Spicer, R. A. & Chapman, J. L. Climate change and the evolution of high-latitude terrestrial vegetation and flora. *Trends Ecol. Evol.* **5**, 279–284 (1990).
- Crowley, T. J. & Berner, R. A. CO₂ and climate change. *Science* **292**, 870–872 (2001).
- Seward, A. C. Antarctic Fossil Plants. British Antarctic ('Terra Nova') Expedition, 1910. *British Museum Natural History Report. Geology* **1**, 1–49 (1914).
- Chaney, R. W. Tertiary centers and migration routes. *Ecol. Monogr.* **17**, 139–148 (1947).
- Hickey, L. J. Eternal summer at 80 degrees north. *Discovery* **17**, 17–23 (1984).
- Wolfe, J. A. in *The Carbon Cycle and Atmospheric CO₂: Natural Variations, Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 357–375 (Geophys. Monogr. Ser. 32, American Geophysical Union, Washington DC, 1985).
- Falcon-Lang, H. J. & Cantrill, D. J. Leaf phenology of some mid-Cretaceous polar forests, Alexander Island, Antarctica. *Geol. Mag.* **138**, 39–52 (2001).
- Estes, R. & Hutchison, J. Eocene lower vertebrates from Ellesmere Island, Canadian Arctic Archipelago. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **30**, 325–347 (1980).
- Tarduno, J. A. *et al.* Evidence for extreme climatic warmth from Late Cretaceous arctic vertebrates. *Science* **282**, 2241–2244 (1998).
- Tripathi, A., Zachos, J., Marinovich, L. & Bice, K. Late Paleocene Arctic coastal climate inferred from molluscan stable and radiogenic isotope ratios. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **170**, 101–113 (2001).
- Dutton, A. L., Lohmann, K. C. & Zinsmeister, W. J. Stable isotope and minor element proxies for Eocene climate of Seymour Island, Antarctica. *Paleoceanography* **17**(2), 1016, doi:10.1029/2000PA000593 (2002).
- Müller, M. J. *Selected Climatic Data for a Global Set of Standard Stations for Vegetation Science* (Kluwer Academic, Dordrecht, The Netherlands, 1981).
- Beerling, D. J. & Osborne, C. P. Physiological ecology of Mesozoic polar forests in a high CO₂ environment. *Ann. Bot.* **89**, 329–339 (2002).
- Osborne, C. P. & Beerling, D. J. A process-based model of conifer forest structure and function with special emphasis on leaf lifespan. *Glob. Biogeochem. Cycles* **16**(4), 1097 doi:10.1029/2001GB001467 (2002).
- Tjoelker, M. G., Oleksyn, J. & Reich, P. B. Modelling respiration of vegetation: evidence for a general

- temperature-dependent Q₁₀. *Glob. Change Biol.* **7**, 223–230 (2001).
- Read, J. & Francis, J. Responses of some Southern Hemisphere tree species to a prolonged dark period and their implications for high-latitude Cretaceous and Tertiary floras. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **99**, 271–290 (1992).
- Villar, R. & Merino, J. Comparison of leaf construction costs in woody species with differing leaf life-spans in contrasting ecosystems. *New Phytol.* **151**, 213–226 (2001).
- Yin, X. Responses of leaf nitrogen concentration and specific leaf area to atmospheric CO₂ enrichment: a retrospective synthesis across 62 species. *Glob. Change Biol.* **8**, 631–642 (2002).
- Valdes, P. J., Sellwood, B. W. & Price, G. D. Evaluating concepts of Cretaceous equability. *Palaeoclim. Data Modell.* **2**, 139–158 (1996).
- Creber, G. T. & Chaloner, W. G. Tree growth in the Mesozoic and Early Tertiary and the reconstruction of palaeoclimates. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **52**, 35–60 (1985).
- Upchurch, G. R. & Askin, R. A. Latest Cretaceous and earliest Tertiary dispersed plant cuticles from Seymour Island. *Antarct. J. US* **24**, 7–10 (1990).
- Parrish, J. T., Daniel, I. L., Kennedy, E. M. & Spicer, R. A. Paleoclimatic significance of mid-Cretaceous floras from the middle Clarence Valley, New Zealand. *Palaios* **13**, 149–159 (1998).
- Axelrod, D. I. Origin of deciduous and evergreen habits in temperate forests. *Evolution* **20**, 1–15 (1966).
- Wolfe, J. A. Late Cretaceous-Cenozoic history of deciduousness and the terminal Cretaceous event. *Paleobiology* **13**, 215–226 (1987).
- Reich, P. B., Walters, M. B. & Ellsworth, D. S. From tropics to tundra: global convergence in plant functioning. *Proc. Natl Acad. Sci. USA* **94**, 13730–13734 (1997).
- Givnish, T. J. Adaptive significance of evergreen vs. deciduous leaves: solving the triple paradox. *Silva Fennica* **36**, 703–743 (2002).
- Falcon-Lang, H. J. The relationship between leaf longevity and growth ring markedness in modern conifer woods and its implications for palaeoclimatic studies. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **160**, 317–328 (2000).
- Beerling, D. J. *et al.* The influence of Carboniferous palaeoatmospheres on plant function: an experimental and modelling assessment. *Phil. Trans. R. Soc. Lond. B* **353**, 131–140 (1998).
- Sokal, R. R. & Rohlf, F. J. *Biometry* (W. H. Freeman, New York, 1995).
- Parton, W. J. *et al.* Observations and modelling of biomass and soil organic matter dynamics for the grassland biome worldwide. *Glob. Biogeochem. Cycles* **7**, 785–809 (1993).

Acknowledgements We thank S. J. Brentnall for running the USCM simulations, P. J. Valdes for providing the Cretaceous GCM climate, W. G. Chaloner, L. J. Hickey, R. J. Norby, G. R. Upchurch, P. Wilf and F. I. Woodward for comments, and the Royal Society (D.J.B. and C.P.O.), the Leverhulme Trust (D.J.B.), the Natural Environmental Research Council, UK (D.J.B.), the US National Science Foundation (D.L.R.) and the US Department of Energy (R. A. Berner) for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.B. (d.j.beerling@sheffield.ac.uk).

Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium

Jianzhong He*, Kirsti M. Ritalahti*, Kun-Lin Yang*, Stephen S. Koenigsberg† & Frank E. Löffler*‡

* School of Civil and Environmental Engineering and

‡ School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30332-0512, USA

† Regenesys Bioremediation Products, 1011 Calle Sombra, San Clemente, California 92672-6244, USA

Tetrachloroethene (PCE) and trichloroethene (TCE) are ideal solvents for numerous applications, and their widespread use makes them prominent groundwater pollutants. Even more troubling, natural biotic and abiotic processes acting on these solvents lead to the accumulation of toxic intermediates (such as dichloroethenes) and carcinogenic intermediates (such as vinyl chloride)^{1–4}. Vinyl chloride was found in at least 496 of the 1,430 National Priorities List sites identified by the US Environmental Protection Agency, and its precursors PCE and TCE are present in at least 771 and 852 of these sites, respectively⁵. Here we describe an unusual, strictly anaerobic bacterium that destroys dichloroethenes and vinyl chloride as part of its energy metabolism,

generating environmentally benign products (biomass, ethene and inorganic chloride). This organism might be useful for cleaning contaminated subsurface environments and restoring drinking-water reservoirs.

Crucial to the detoxification of chloroethene-contaminated sites is the complete reductive dechlorination of these contaminants to non-chlorinated end products such as ethene, or their oxidation to carbon dioxide (mineralization). Microbial growth linked to the mineralization of *cis*-dichloroethene (*cis*-DCE) and vinyl chloride (VC) occurs under aerobic conditions^{6,7}, but VC is generated frequently from polychlorinated ethenes in anoxic and reduced environments. Thus, an anaerobic process that leads to complete detoxification would be most effective in achieving bioremediation *in situ*. Substantial information describing bacteria that use polychlorinated ethenes as metabolic electron acceptors has accumulated^{8,9} but the populations capable of complete reductive dechlorination have remained elusive. *Dehalococcoides ethenogenes* strain 195 (GenBank accession number AF004928.2) has been shown to dechlorinate PCE to ethene, but this organism failed to gain energy from VC, slowly producing ethene in a co-metabolic process¹⁰. To exploit the reductive dechlorination process for environmental cleanup, finding organisms that efficiently reduce VC to ethene is a priority.

Starting with a microcosm capable of dechlorinating PCE to ethene¹¹, we isolated a novel bacterium that used VC as a growth-supporting electron acceptor, thereby transforming this carcinogenic compound to the benign products ethene and inorganic chloride. The isolation procedure took advantage of the organism's ability to derive all its energy required for growth from the reduction of VC to ethene. Continued transfers over 4 years in mineral salts medium amended with VC, hydrogen and acetate yielded a non-methanogenic, ethene-producing culture. Dechlorination occurred with acetate as the sole electron donor, although at lower rates, apparently mediated in association with a syntrophic, acetate-oxidizing partner population^{11,12}. Consecutive transfers

without hydrogen achieved further enrichment of the dechlorinating population. VC dechlorination activity was recovered repeatedly from 10^{-5} dilutions of consecutive dilution-to-extinction series in hydrogen-amended medium. Microscopic examination after this enrichment procedure revealed the presence of three morphotypes: a small, disc-shaped organism and two rod-shaped organisms, one short and one long. Early attempts to tease out the VC-dechlorinating population in pure culture by using the dilution-to-extinction principle as well as cultivation in semisolid medium containing 0.5% low-melting agarose were unsuccessful. Similarly to *Dehalococcoides* strains 195 (ref. 10) and CBDB1 (ref. 13), the addition of high concentrations of the peptidoglycan inhibitor ampicillin did not prohibit dechlorination, and after five consecutive transfers in liquid medium with 1 mg ml^{-1} ampicillin, the rod-shaped organisms were no longer detectable by microscopic examination. After treatment with ampicillin, dechlorinating activity was recovered repeatedly from 10^{-7} dilutions in defined basal salts medium amended with VC, hydrogen and acetate. In addition, VC dechlorination occurred after transferring tiny opaque colonies that developed after 4–5 weeks in semisolid medium to liquid medium. No growth occurred in complex medium, and the culture seemed microscopically homogeneous (Fig. 1a). The isolate, designated BAV1, is a small non-motile organism of no more than $0.8 \mu\text{m}$ in diameter (Fig. 1). Careful light-microscopic analysis suggested that BAV1 cells are disc-shaped rather than coccoid, supporting an earlier observation indicating that *Dhc. ethenogenes* strain 195 cells might be flattened¹⁰. Cells suspended in liquid seemed to tumble end over end, a phenomenon explained by a disc-shaped morphology. High-resolution scanning electron microscopy also suggested that BAV1 cells have a disc-shaped rather than a coccoid morphology (Fig. 1b). Figure 1b,c shows individual BAV1 cells and features peculiar filamentous appendages on the cell's surface. Different 16S ribosomal RNA gene-based approaches corroborated

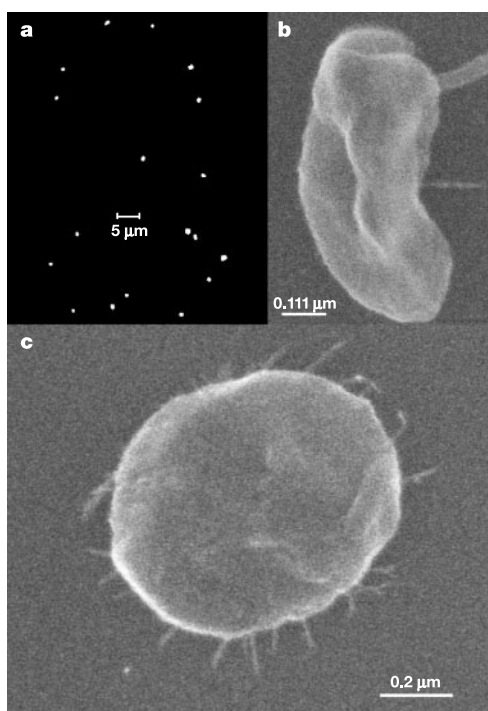


Figure 1 Micrographs of isolate BAV1. **a**, Epifluorescence observed after acridine orange staining. Original magnification $\times 1,000$. **b**, **c**, Scanning electron micrographs indicating a disc-shaped morphology (**b**) and displaying peculiar appendages (**c**).

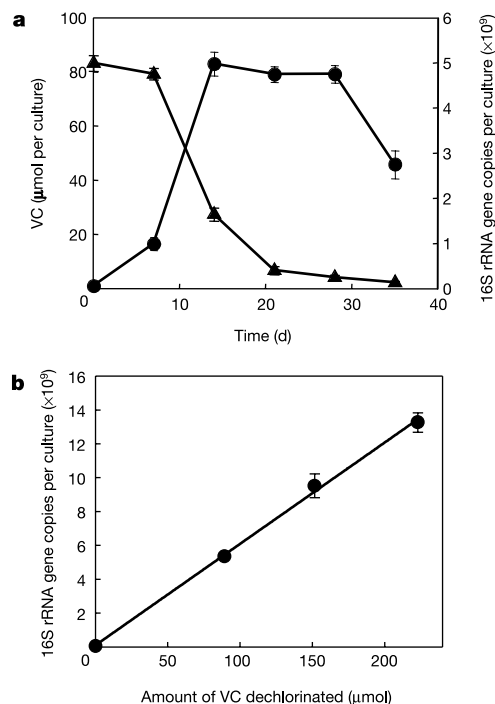


Figure 2 VC-dependent growth of isolate BAV1. **a**, An increase in 16S rRNA gene copies as determined by real-time PCR (circles) during the reductive dechlorination of VC (triangles) to ethene (not shown) by culture BAV1. **b**, 16S rRNA gene copies of isolate BAV1 after dechlorinating different amounts of VC. Data points were averaged from triplicates; error bars $\pm$ s.d. are not shown if they are hidden by the symbol.

Table 1 Electron acceptor utilization profiles of defined *Dehalococcoides*-like populations

<i>Dehalococcoides</i> sp.*	Reference	Electron acceptors
<i>Dhc. ethenogenes</i> strain 195 (C, AF004928.2)	10	PCE, TCE, <i>cis</i> -DCE, 1,1-DCE, 1,2-dichloroethane, 1,2-dibromoethane
<i>Dhc.</i> sp. strain VS (V, AF388550)†	15	<i>cis</i> -DCE, VC
<i>Dhc.</i> sp. strain FL2 (P, AF357918.2)	9	TCE, <i>cis</i> -DCE
<i>Dhc.</i> sp. strain BAV1 (P, AY165308)	This study	<i>cis</i> -DCE, <i>trans</i> -DCE, 1,1-DCE, VC, vinyl bromide, 1,2-dichloroethane
<i>Dhc.</i> sp. strain CBDB1 (P, AF230641)	13,25	1,2,3-TCB, 1,2,4-TCB, 1,2,3,4-TeCB, 1,2,3,5-TeCB, 1,2,4,5-TeCB, PCDD†

*Group designations according to ref. 14: C, Cornell; V, Victoria; P, Pinellas sequence subgroups; also indicated are the GenBank accession numbers and references.

†Bacterium VS was identified in a mixed culture.

‡TCB, trichlorobenzene; TeCB, tetrachlorobenzene; PCDD, polychlorinated dibenzo-*p*-dioxins.

the purity of the culture. Terminal restriction-fragment-length polymorphism revealed single peaks of the expected sizes of 197, 442 and 512 base pairs (bp) after digestion of the amplicons with the restriction enzymes *Hha*I, *Msp*I and *Rsa*I, respectively. Denaturing gradient gel electrophoresis analysis yielded a single 148-bp band, whose sequence exactly matched that of isolate BAV1. Amplified ribosomal DNA restriction analysis (ARDRA) of 80 clones from two 16S rRNA gene clone libraries established with genomic DNA obtained from a VC-grown culture generated patterns predicted by digestion *in silico*, indicating that all 16S rRNA gene inserts contained in the clone libraries belonged to isolate BAV1. The fact that the *Dhc. ethenogenes* strain 195 genome (www.tigr.org) possesses a single ribosomal RNA operon suggested that a particular 16S rRNA gene-based analytical technique would not be complicated by the possible variations in multiple (slightly different) 16S rRNA gene sequences.

Isolate BAV1 respired VC in defined, completely synthetic basal salts medium amended with acetate and hydrogen (Fig. 2a). At room temperature (22–25 °C), BAV1 dechlorinated VC at up to $134.2 \pm 10 \text{ nmol min}^{-1}$ per mg of protein, and grew with a doubling time of 2.2 days to yield $239 \pm 27 \text{ mg}$ (mean $\pm$ s.d., $n = 6$) of protein per mole of chloride released. Growth depended strictly on reductive dechlorination and the presence of hydrogen as an electron donor, which could not be replaced by organic substrates including formate, acetate, lactate, pyruvate, propionate, glucose, ethanol or yeast extract. Besides VC, other growth-supporting electron acceptors included *cis*-DCE, *trans*-DCE, 1,1-DCE or 1,2-dichloroethane and vinyl bromide; stoichiometric amounts of ethene accumulated as the reduced end product. It is notable that BAV1 is the first isolate capable of the metabolic dechlorination of all DCE isomers. Chlorinated compounds not supporting growth included PCE, TCE, chlorinated propanes, 1,1,1-trichloroethane, 1,1-dichloroethane and chloroethane. However, PCE and TCE were co-metabolized in the presence of a growth-supporting chloroethene and ethene was produced. Other organic and inorganic electron acceptors such as nitrate, fumarate, ferric iron, sulphite, sulphate, thiosulphate, sulphur or oxygen were not utilized, and no fermentative growth was observed. Respiratory growth was demonstrated conclusively by the chloroethene-dependent increase in cellular macromolecules (such as protein and DNA). Neither cell proliferation nor protein increase was detected in cultures lacking VC, acetate or hydrogen. Real-time polymerase chain reaction (PCR) showed that the increase in cell numbers was concomitant with the consumption of VC (Fig. 2a). A linear increase in biomass (that is, cells, as measured by the increase in 16S rRNA gene copies) occurred with increasing amounts of VC provided as electron acceptor, indicating a tight coupling between reductive dechlorination and growth (Fig. 2b). The number of 16S rRNA gene copies measured in cultures without VC corresponded to the number of cells transferred with the inoculum, confirming that no growth occurred in the absence of VC. Cultures that had consumed $80 \mu\text{mol}$ of *cis*-DCE contained about twice as many 16S rRNA gene copies than cultures grown with $80 \mu\text{mol}$ of VC (for example, $(9.28 \pm 0.41) \times 10^9$ copies versus $(4.99 \pm 0.26) \times 10^9$ copies).

These findings show that BAV1 captured energy from both dechlorination steps when grown with *cis*-DCE.

Phylogenetic analysis, performed by using double-stranded 16S rRNA gene sequencing, affiliated isolate BAV1 (AY165308) with the Pinellas group of the *Dehalococcoides* cluster¹⁴, a deep branch within the phylum Chloroflexi (green non-sulphur bacteria)¹³. The Pinellas group also includes *Dehalococcoides* sp. strains CBDB1 (AF230641)¹³ and FL2 (AF357918.2)⁹. Metabolic ethene formation is not restricted to members of the Pinellas group and was described in a mixed culture containing a *Dehalococcoides* population of the Victoria group¹⁵. Table 1 shows known metabolic electron acceptors along with the phylogenetic grouping of identified *Dehalococcoides*-like populations. BAV1 shares a highly similar 16S rRNA gene sequence with *Dehalococcoides* populations that failed to dechlorinate chloroethenes or grow with VC as a metabolic electron acceptor (for example, strains CBDB1 and FL2). This high degree of 16S rRNA gene sequence similarity among members of the *Dehalococcoides* cluster implies that an analysis of 16S rRNA gene sequences cannot distinguish between populations of this group exhibiting different physiological activities. Hence, focusing only on 16S rRNA gene sequence analysis is not sufficient for characterizing the dechlorinating community or for the reliable prediction of the dechlorination potential associated with a particular environment.

Complete reductive dechlorination of chlorinated ethenes to non-toxic end products has been documented extensively in laboratory studies^{12,15,16}, and ethene formation was observed at some contaminated sites¹⁴. However, until now no organisms that efficiently dechlorinate VC have been obtained in pure culture. The isolation of the VC-respiring population BAV1 is a relevant milestone in chloroethene detoxification and provides deeper insight into poorly understood halogen cycles that involve the natural formation and utilization of chloroorganic compounds, including VC^{17,18}. The comprehension of such cycles might explain why complete reductive dechlorination does not occur at all contaminated sites, even after alteration and optimization of geochemical conditions or supplying suitable electron donors to accelerate microbial activity. A recent pilot demonstration in a hydraulically controlled recirculation test plot at the chloroethene-contaminated Bachman Road site in Oscoda, Michigan, supports bioaugmentation with BAV1 as a promising approach to achieve detoxification at sites where the indigenous microbiota to drive the reductive dechlorination process to completion is absent or the rates of contaminant removal are insufficient¹⁹. Such innovative technologies are needed to clean up numerous chloroethene-contaminated aquifers at reasonable costs within acceptable time frames, and to protect threatened drinking-water reservoirs. □

Methods

Medium preparation

Completely defined, anaerobic mineral salts medium was prepared as described^{12,20,21}. The enrichment and isolation process was performed in 20-ml glass vials containing 10 ml (final volume) of growth medium, whereas the cultures used for kinetic studies or molecular analyses were grown in 160-ml serum bottles containing 100 ml of medium. Unless indicated otherwise, cultures were incubated at 30 °C in the dark without shaking. Soluble substrates were added at 5 mM. Hydrogen was added by syringe to at least double

the concentration required for complete reductive dechlorination (namely 8.5–17 kPa). Chlorinated compounds were added to final aqueous concentrations in the range of 0.47–1.33 mM. Semisolid medium was prepared by adding 0.5% w/w low-melting agarose before autoclaving.

Analytical techniques

The protein content of liquid cultures was estimated as follows. Cells were harvested from 8 ml of culture fluid by centrifugation (10 min, 10,000 g). After alkaline cell lysis²², the Coomassie Plus Protein Assay Reagent Kit (Pierce Biotechnology, Rockford, Illinois) was used in accordance with the manufacturer's recommendations. Spectrophotometric determination at 595 nm was used to quantify protein by comparing the sample absorbance with protein standards of known concentration prepared in the same way as the samples. Chloroethene and ethene concentrations were determined by gas chromatography as described¹².

Electron microscopy

A Zeiss LSM 510 confocal microscope with a Plan-Neofluar objective (100×; numerical aperture, 1.3) was used to obtain micrographs of cell suspensions after staining with acridine orange (15–60 min in 0.01% aqueous solution). Scanning electron microscopy micrographs were obtained with a TOPCON DS-130 field-emission scanning electron microscope with samples staged 'in-lens' and photographed at 20 kV. Samples were prepared as described²³, and coated with chromium (about 1.5 nm thickness) in a Denton DV-602 turbo magnetron sputter system.

Molecular analyses

DNA was extracted from actively growing cultures as described¹². Real-time PCR to quantify BAV1 cells used a probe targeted to the *Dehalococcoides* 16S rRNA gene, tagged with a 6-carboxyfluorescein reporter fluorochrome on the 5' end, and *N,N,N',N'*-tetramethyl-6-carboxyrhodamine quencher on the 3' end as described previously¹². Linear calibration curves ($r^2 > 0.99$) were generated, spanning a template concentration range from 6.9×10^2 to 6.9×10^6 16S rRNA gene copies per 30-μl reaction volume by using BAV1 genomic DNA or plasmid DNA containing the 16S rRNA gene from BAV1. Analysis of individual clones with ARDRA was performed as described previously^{12,24}, except that the PCR-amplified 16S rRNA gene products from each clone were digested for 3 h with enzymes *HhaI*, *MspI* and *RsaI* at 37 °C. The reactions were terminated by incubation at 65 °C for 10 min, in accordance with the manufacturer's recommendations (Gibco), and the resulting fragments were resolved by electrophoresis for 2 h on 2.5% low-melting agarose gel (Seaplaque; Cambrex, Rockland, Maine). The fluorescently labelled primer 8F-hex (5'-AGA GTT TGA TCC TGG CTC AG-3') and unlabelled 1492R (5'-GC(C/T) TAC CTT GTT ACG ACT T-3') were used to amplify the 16S rRNA gene from pure culture DNA. Fluorescently labelled terminal fragments obtained by digesting the PCR product with *HhaI*, *MspI* and *RsaI* were analysed at Michigan State University's Genomics Technology Support Facility. Denaturing gradient gel electrophoresis was performed by Microbial Insights (Rockford, Tennessee) with the use of universal bacterial primers corresponding to *Escherichia coli* positions 341–534 as described¹⁶.

Received 29 January; accepted 29 April 2003; doi:10.1038/nature01717.

- Kielhorn, J., Melber, C., Wahnschaffe, U., Aitio, A. & Mangelsdorf, I. Vinyl chloride: still a cause for concern. *Environ. Health Perspect.* **108**, 579–588 (2000).
- Mohn, W. W. & Tiedje, J. M. Microbial reductive dehalogenation. *Microbiol. Rev.* **56**, 482–507 (1992).
- Roberts, A. L. *et al.* Reductive elimination of chlorinated ethylenes by zero-valent metals. *Environ. Sci. Technol.* **30**, 2654–2659 (1996).
- Vogel, T. M., Criddle, C. S. & McCarty, P. L. Transformation of halogenated aliphatic compounds. *Environ. Sci. Technol.* **21**, 722–736 (1987).
- US Environmental Protection Agency. Agency for Toxic Substances and Disease Registry, ToxFQs for chlorinated ethenes; www.atsdr.cdc.gov/tfacts70.html (1996).
- Coleman, N. V., Mattes, T. E., Gossett, J. M. & Spain, J. C. Phylogenetic and kinetic diversity of aerobic vinyl chloride-assimilating bacteria from contaminated sites. *Appl. Environ. Microbiol.* **68**, 6162–6171 (2002).
- Coleman, N. V., Mattes, T. E., Gossett, J. M. & Spain, J. C. Biodegradation of *cis*-dichloroethene as the sole carbon source by a β -Proteobacterium. *Appl. Environ. Microbiol.* **68**, 2726–2730 (2002).
- Holliger, C., Wohlfarth, G. & Diekert, G. Reductive dechlorination in the energy metabolism of anaerobic bacteria. *FEMS Microbiol. Rev.* **22**, 383–398 (1998).
- Löffler, F. E., Cole, J. R., Ritalahti, K. M. & Tiedje, J. M. in *Dehalogenation: Microbial Processes and Environmental Applications* (eds Häggblom, M. M. & Bossert, I. D.) 53–87 (Kluwer Academic, New York, 2003).
- Maymó-Gatell, X., Chien, Y.-T., Gossett, J. M. & Zinder, S. H. Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* **276**, 1568–1571 (1997).
- He, J. *et al.* Acetate versus hydrogen as direct electron donors to stimulate the microbial reductive dechlorination process at chloroethene-contaminated sites. *Environ. Sci. Technol.* **36**, 3945–3952 (2002).
- He, J., Ritalahti, K. M., Aiello, M. R. & Löffler, F. E. Complete detoxification of vinyl chloride (VC) by an anaerobic enrichment culture and identification of the reductively dechlorinating population as a *Dehalococcoides* population. *Appl. Environ. Microbiol.* **69**, 996–1003 (2003).
- Adrian, L., Szwedzyk, U., Wecke, J. & Görsch, H. Bacterial dehalorespiration with chlorinated benzenes. *Nature* **408**, 580–583 (2000).
- Hendrickson, E. R. *et al.* Molecular analysis of *Dehalococcoides* 16S ribosomal DNA from chloroethene-contaminated sites throughout North America and Europe. *Appl. Environ. Microbiol.* **68**, 485–495 (2002).
- Cupples, A. M., Spormann, A. M. & McCarty, P. L. Growth of a *Dehalococcoides*-like microorganism on vinyl chloride and *cis*-dichloroethene as electron acceptors as determined by competitive PCR. *Appl. Environ. Microbiol.* **69**, 953–959 (2003).
- Duhamel, M. *et al.* Comparison of anaerobic dechlorinating enrichment cultures maintained on

- tetrachloroethene, trichloroethene, *cis*-dichloroethene and vinyl chloride. *Water Res.* **36**, 4193–4202 (2002).
- Gribble, G. W. in *Chlorinated Compounds in the Biosphere, Natural Production* (ed. Meyers, R. A.) 972–1035 (Wiley, New York, 1998).
- Keppeler, F., Borchers, R., Pracht, J., Rheinberger, S. & Scholer, H. F. Natural formation of vinyl chloride in the terrestrial environment. *Environ. Sci. Technol.* **36**, 2479–2483 (2002).
- Lendvay, J. M. *et al.* Bioreactive barriers: bioaugmentation and biostimulation for chlorinated solvent remediation. *Environ. Sci. Technol.* **37**, 1422–1431 (2003).
- Löffler, F. E., Ritalahti, K. M. & Tiedje, J. M. Dechlorination of chloroethenes is inhibited by 2-bromoethanesulfonate in the absence of methanogens. *Appl. Environ. Microbiol.* **63**, 4982–4985 (1997).
- Löffler, F. E., Tiedje, J. M. & Sanford, R. A. Fraction of electrons consumed in electron acceptor reduction and hydrogen threshold as indicators of halo-respiratory physiology. *Appl. Environ. Microbiol.* **65**, 4049–4056 (1999).
- Gerhardt, P. (ed.) *Manual of Methods for General Bacteriology* (American Society for Microbiology, Washington DC, 1981).
- Sung, Y. *et al.* Characterization of two tetrachloroethene-reducing, acetate-oxidizing anaerobic bacteria and their description as *Desulfuromonas michiganensis* sp. nov. *Appl. Environ. Microbiol.* **69**, 2964–2974 (2003).
- Löffler, F. E., Sun, Q., Li, J. & Tiedje, J. M. 16S rRNA gene-based detection of tetrachloroethene (PCE)-dechlorinating *Desulfuromonas* and *Dehalococcoides* species. *Appl. Environ. Microbiol.* **66**, 1369–1374 (2000).
- Bunge, M. *et al.* Reductive dehalogenation of chlorinated dioxins by an anaerobic bacterium. *Nature* **421**, 357–360 (2003).

Acknowledgements Electron micrographs were obtained by R. P. Apkarian at the Integrated Microscopy and Microanalytical Facility at Emory University, Atlanta, Georgia. This work was supported by the Strategic Environmental Research and Development Program, and by a National Science Foundation CAREER award to F.E.L.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.E.L. (frank.loeffler@ce.gatech.edu).

Delta-wing function of webbed feet gives hydrodynamic lift for swimming propulsion in birds

L. Christoffer Johansson* & R. Åke Norberg†

* Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA

† Department of Zoology, Göteborg University, Box 463, SE-405 30 Göteborg, Sweden

Most foot-propelled swimming birds sweep their webbed feet backwards in a curved path that lies in a plane aligned with the swimming direction. When the foot passes the most outward position, near the beginning of the power stroke, a tangent to the foot trajectory is parallel with the line of swimming and the foot web is perpendicular to it. But later in the stroke the foot takes an increasingly transverse direction, swinging towards the longitudinal axis of the body. Here we show that, early in the power stroke, propulsion is achieved mostly by hydrodynamic drag on the foot, whereas there is a gradual transition into lift-based propulsion later in the stroke. At the shift to lift mode, the attached vortices of the drag-based phase turn into a starting vortex, shed at the trailing edge, and into spiralling leading-edge vortices along the sides of the foot. Because of their delta shape, webbed feet can generate propulsive forces continuously through two successive modes, from drag at the beginning of the stroke, all the way through the transition to predominantly lift later in the stroke.

Foot propulsion in swimming birds has generally been considered to be drag-based, both in surface swimming and during diving^{1,2}. Drag is the hydrodynamic force that opposes movement

through a fluid; it is aligned with the movement but acts in the opposite direction. This mode requires that the foot is moved backwards in a still-water frame of reference. The foot must therefore move backwards faster, relative to the bird, than it swims. An entirely different mode of foot-propelled swimming has recently been detected in the great crested grebe (*Podiceps cristatus*)^{3,4}. It does not paddle with its feet in a fore-and-aft movement, but instead sweeps its feet outwards-upwards in a transverse plane, perpendicular to the swimming direction. There is no backward movement of the feet relative to the water, so drag cannot be tilted forwards and so is useless for propulsion. Instead, propulsion is based on lift, which here is directed forwards and is nearly perpendicular to the path of the foot^{3,4}, being by definition normal to the resultant relative flow. There are two functional benefits from using lift rather than pure drag for propulsion. First, with drag-based propulsion, muscle work is done directly against the entire reaction force on the foot, whereas work is done against only a component of it when propulsion is based on lift. And that component is often smaller than the propulsive force. In lift mode it is therefore possible to exploit a propulsive force that is larger than the force against which work is being done. This enhances hydrodynamic efficiency and energy economy. Second, with drag-based propulsion the feet must move backwards in still water, so the bird must sweep its feet faster than it swims. With lift-based propulsion it need not do so, which again improves energy economy and also might allow higher swimming speeds because the sweep velocity of the foot in lift mode need not be nearly as high as the swimming speed.

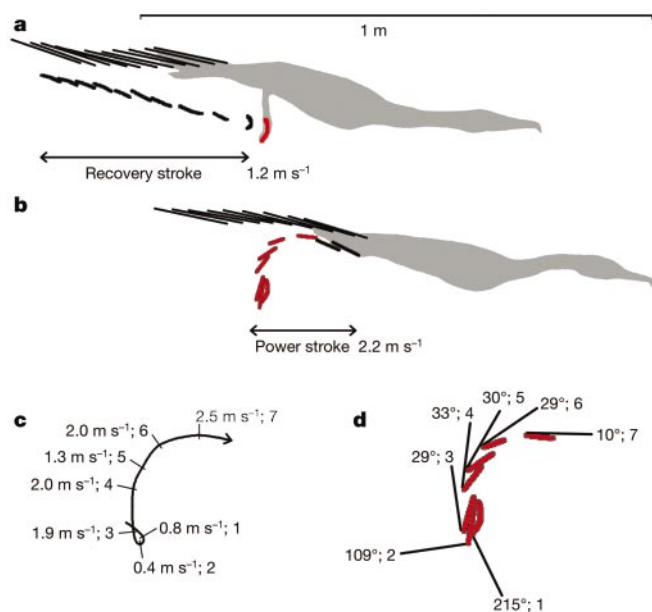


Figure 1 Tracings from a video sequence (separated by 0.02 s) of a foot-stroke cycle of a diving cormorant. Bold lines indicate the position and orientation of the foot, and thin lines show the dorsal side of the tail. **a**, The recovery stroke with the shaded bird at the start of the power stroke. **b**, The power stroke, starting with the same foot position that terminated the upper sequence. The shaded bird is at the end of the power stroke. Acceleration of the body occurs over seven video fields (from 1.2 m s⁻¹ at the end of recovery stroke to 2.2 m s⁻¹ at position 7 of the power stroke), indicated by the red foot. **c**, **d**, Data for the acceleration phase of **b**. **c**, The instantaneous still-water speed of the trailing edge of the foot is indicated at positions 1–7 along its trajectory. **d**, The instantaneous angle of attack at the trailing edge of the foot is indicated at the same seven foot positions. The drag mode can last, at most, for as long as the foot moves backwards relative to the still water, which it does until slightly after position 3, during which time the angle of attack changes from ~215° to less than 30°. This is 39% of the duration of the power stroke. During the remainder of the power stroke, 61% of the time, the body continues to accelerate forwards, owing entirely to lift.

Here we show that the more conventional fore-and-aft movement of the feet in swimming birds can also generate thrust that is based on hydrodynamic lift. Our results are based on observations and analyses of video films of swimming and diving birds (Fig. 1), and on flow-visualization experiments with a model foot in a simulated power stroke (Fig. 2).

We distinguish three phases of the power stroke. Phase 1 is when the foot performs a largely translatable backward movement near the start of the power stroke. A recirculating wake is formed on the leeward, suction side of the foot, and a closed vortex, surrounding the foot, is rolled up behind its edges. At this stage, the foot moves backwards relative to the water and water meets the foot at nearly right angles to its web. Water is accelerated and drawn backwards along with the foot, over its leeward surface, and the induced flow is bounded by an attached closed vortex. The reaction force on the foot is hydrodynamic drag, directed forwards, opposite to the direction of movement of the foot. It is the useful propulsive force at this stage (Fig. 3).

Because of rotation of the leg, the outer, distal edge of the foot moves faster than its proximal parts, and this velocity gradient creates a stronger suction at the distal edge and a lower pressure within the vortex core there. Water should therefore be drawn towards the foot's trailing edge on the suction side and spiral rearwards in the side-edge vortices, similar to the spanwise, leading-edge flow over flapping insect wings⁵. In this way a rearward flow is set up on the suction side, typical for a wing in lift mode.

Stage 2 of the power stroke occurs after the foot has passed its most outward position: the path curves inwards, towards the longitudinal body axis of the bird. The transverse speed component therefore increases gradually while the backward speed component diminishes. The backward speed component very soon becomes lower than the bird's swimming speed, so the foot no longer moves backwards relative to the water but instead moves forwards with the bird. When this happens, the drag force cannot be tilted forwards, so propulsion is no longer drag-based (Figs 1 and 3).

When the foot starts moving transversely in relation to the swimming direction, it leaves behind the previous, attached, trailing-edge vortex, which now becomes a starting vortex for a drawn-out, U-shaped vortex behind the foot. The vortices that were attached to the sides of the foot in drag mode turn into spiralling leading-edge vortices, which leave stable, trailing tip-vortices behind the outer corners of the foot, as on delta-wings^{6,7}. They form the sides of the shed U-shaped vortex (Figs 2 and 3).

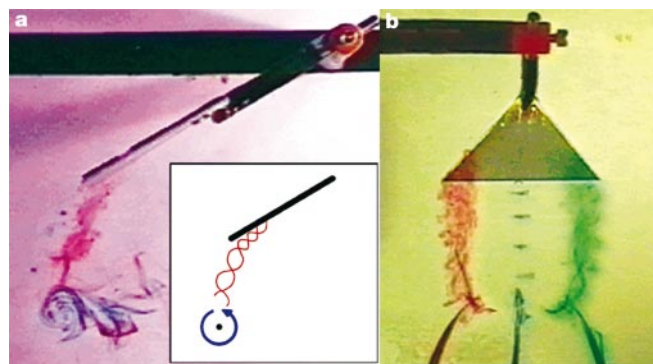


Figure 2 Side and front views of the model foot simulating a power stroke of a swimming bird. **a**, Side view. The foot rotates in a clockwise direction and also translates to the right. **b**, Front view, showing the leeward, suction, side of the foot as seen from the right side in **a**. The flow swirls around the swept-back side edges of the model foot, over to its suction side, where it is entrained in an attached leading-edge vortex that is subsequently shed from the outer corner of the foot to form a spiralling, trailing tip-vortex. The side view also shows a transverse starting vortex left behind in the wake when the model foot moves upwards and to the right.

Circulation is in the same sense as when the vortices were attached to the foot when it translated backwards, in drag mode. Enclosed by the U-shaped vortex is a jet of water flowing backwards at about right angles to the vortex plane. In addition the vortex itself is convected downstream, but at a slower speed than the induced jet (Fig. 3).

Lift production should be facilitated early in the lift phase because there is probably a rearward flow along the suction side of the foot already in drag mode. In addition, the trailing-edge vortex—also developed in drag mode—is just shed as a starting vortex at the transition to lift mode, and the attached side-edge vortices are also built up in drag mode and later turn directly into leading-edge vortices in the lift phase.

When the relative water velocity due to the swimming speed is added vectorially to the velocity due to the foot movement relative to the bird, tangential to the stroke path, the resultant relative water speed is seen to meet the foot from obliquely in front, forming an acute angle of attack with the foot (Fig. 3). This angle depends on the position of the foot in its path and on the relation between the swimming speed and the stroke speed, variables that change throughout the power stroke and along the foot.

The reaction force on the foot is directed obliquely forwards, and is opposite to the direction of the water flow in the jet enclosed by

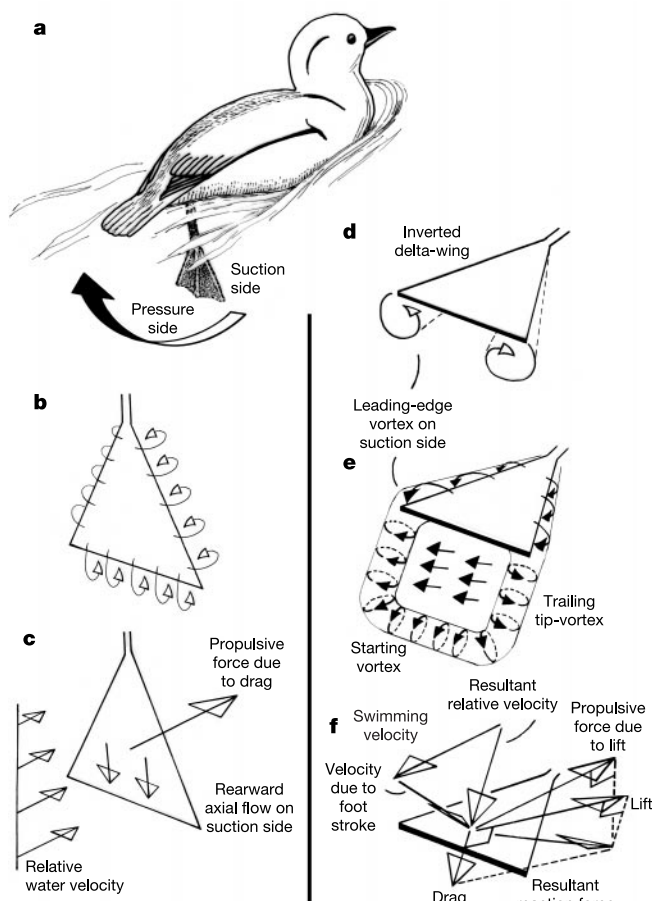


Figure 3 Schematic, generalized representation of flow patterns and forces set up by a webbed foot during a power stroke of a bird in fast swimming. **a**, Definition of suction side and pressure side of the foot in backward motion. The elicited forces are inferred from the induced water flow, as observed in our flow-visualization experiments. **b**, **c**, The foot in drag mode; **d**–**f**, the foot in lift mode at a later stage of the power stroke. The ratio of swimming velocity to foot-stroke velocity used for **f** is roughly as in lift mode at foot position 6 in Fig. 1, and the lift-to-drag ratio is just an example for one phase of the stroke. All velocity and force vectors lie in a common plane aligned with the line of swimming.

the vortex. It is only because of the production of a lift force that the resultant reaction force gets a forward inclination, owing to which there is a forward force vector. This is the useful propulsive force and it is due entirely to lift. Drag forces on the foot only detract from the propulsive force at this stage (Fig. 3).

In phase 3, at the end of the power stroke the foot is furred and pulled out of the wake and the U-shaped vortex closes after the foot to form an ellipse.

The motion of the model foot in this study is the simple consequence of adding a rotary, sweeping, motion of the foot to the forward velocity of the bird. This suggests that the mechanism described here might also be exploited by other swimming animals with webbed feet or fins. The Δ form occurs in the caudal fins of arrow-worms (phylum Chaetognatha), the caudal and pectoral fins of fishes, in the webbed feet of frogs and most swimming birds, the hind flippers of seals, and the tail flukes of whales and dugongs. In addition, the tail of many birds attains a perfect Δ form when activated and spread, owing to its Δ -notched form when furred. Its delta-wing function has recently been treated theoretically⁸ and verified on swallow tails⁹. Hydrofoils and airfoils with a Δ form have thus originated repeatedly and independently by striking evolutionary convergence among widely separate phyletic lines.

A characteristic of a delta-wing is that it is almost un stallable and continues to give large lift forces even at very large angles of attack, although at the penalty of high drag^{6,7}. The fluid dynamics of the Δ form is also insensitive to differences in size and speed, as reflected in the Reynolds number⁷. These features enable a Δ -shaped foot or fin to produce large propulsive forces during the entire power stroke when it operates in two successive modes, from pure drag in the beginning of the stroke, all the way through the transition to predominantly lift in later stages. □

Methods

The kinematics of a diving great cormorant (*Phalacrocorax carbo*) was determined by an analysis of video sequences resolved into 50 fields s^{-1} . The velocities of the bird and the feet were estimated by applying a cubic spline fit¹⁰ to the kinematic data. Angles of attack at the distal end of the feet were determined as the angle between the instantaneous, local velocity vector and the foot as seen in lateral view (Fig. 1).

We swept a model foot through water in a movement pattern closely mimicking the natural movement of the diving cormorant. The foot was made of a 2-mm-thick sheet of aluminium, had an isosceles triangular planform with lateral sides 100 mm long, a distal side 85 mm long and a sweepback angle of 65°. This is a simplified, generalized planform for the model foot that is meant to represent the webbed feet of most swimming birds (even though the cormorant foot has a skewed, triangular planform). The paddling movement of the foot was achieved by using a device consisting of two wheels (250 mm diameter) mounted 150 mm apart on a common horizontal axis, which extended 200 mm on the outer side of one of the wheels. The axis was fixed so that it rotated with the wheels. A thin bar was fastened at 90° to the sticking-out free end of the axis, like a single spoke, and the model foot was attached with its leading apex to the distal end of this transverse bar. The entire device was placed on a horizontal ramp inside a water tank. Ramp elevation was adjusted so that the foot became submerged to a suitable depth in water. By adjusting the length of the transverse bar in relation to the radius of the wheels, it was possible to mimic the natural foot movement in a power stroke. The wheels were rolled manually on the ramp as required to simulate a power stroke. The forward movement of the wheel axis corresponded to the bird's swimming speed, and rotation of the axis with the attached model simulated the foot sweep. This combination of a translatory, forward motion of the body and a rotary, backward sweep of the foot characterizes the kinematics of most swimming birds that use a backward power stroke confined to a plane aligned with the line of swimming. We revealed the induced water flow with a dye ejected via plastic tubes with an internal diameter of 1.14 mm; the tubes were attached along the oblique leading edge on the pressure side of the model foot. The flow thus revealed was examined on video films (Fig. 2).

Received 23 December 2002; accepted 22 April 2003; doi:10.1038/nature01695.

- Blake, R. W. Mechanics of drag-based mechanisms of propulsion in aquatic vertebrates. *Symp. Zool. Soc. Lond.* **48**, 29–52 (1981).
- Braun, J. & Reif, W.-E. A survey of aquatic locomotion in fishes and tetrapods. *N. Jb. Geol. Paläont. Abh.* **169**, 307–332 (1985).
- Johansson, L. C. & Lindhe Norberg, U. M. Asymmetric toes aid underwater swimming. *Nature* **407**, 582–583 (2000).
- Johansson, L. C. & Lindhe Norberg, U. M. Lift-based paddling in diving grebe. *J. Exp. Biol.* **204**, 1687–1696 (2001).
- Ellington, C. P., van den Berg, C., Willmott, A. P. & Thomas, A. L. R. Leading-edge vortices in insect flight. *Nature* **384**, 626–630 (1996).
- Katz, J. & Plotkin, A. *Low-speed Aerodynamics* (McGraw-Hill, New York, 1991).

7. Lee, M. & Ho, C.-M. Lift force of delta wings. *Appl. Mech. Rev.* **43**, 209–221 (1990).
8. Thomas, A. L. R. On the aerodynamics of birds' tails. *Phil. Trans. R. Soc. Lond. B* **340**, 361–380 (1993).
9. Norberg, R. Å. Swallow tail streamer is a mechanical device for self-deflection of tail leading edge, enhancing aerodynamic efficiency and flight manoeuvrability. *Proc. R. Soc. Lond. B* **257**, 227–233 (1994).
10. Walker, J. A. QuickSAND. Quick Smoothing and Numerical Differentiation for the Power Macintosh (<http://www.usm.maine.edu/~walker/software.html>) (1997).

Acknowledgements We thank B. Svensson for building the mechanical experimental device.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.Å.N. (ake.norberg@zool.gu.se).

Chemical and behavioural characterization of the rabbit mammary pheromone

Benoist Schaal*†, Gérard Coureaud*†, Dominique Langlois†‡, Christian Giniès‡, Etienne Sémon‡ & Guy Perrier§

* Centre Européen des Sciences du Goût, CNRS (fre 2328), 21000 Dijon, France

† Unité Mixte de Recherche sur les Arômes, Inra, 21000 Dijon, France

§ Établissement National d'Enseignement Supérieur Agricole, 21000 Dijon, France

† These authors contributed equally to this work

Mammals owe part of their evolutionary success to the harmonious exchanges of information, energy and immunity between females and their offspring. This functional reciprocity is vital for the survival and normal development of infants, and for the inclusive fitness of parents^{1,2}. It is best seen in the intense exchanges taking place around the mother's offering of, and the infant's quest for, milk. All mammalian females have evolved behavioural and sensory methods of stimulating and guiding their inexperienced newborns to their mammae, whereas newborns have coevolved means to respond to them efficiently³. Among these cues, maternal odours have repeatedly been shown to be involved^{4–6}, but the chemical identity and pheromonal nature of these cues have not been definitively characterized until now. Here we focus on the nature of an odour signal emitted by the female rabbit to which newborn pups respond by attraction and oral grasping, and provide a complete chemical and behavioural description of a pheromone of mammary origin in a mammalian species.

The extremely parsimonious strategy of maternal care evolved by the rabbit, *Oryctolagus cuniculus*, offers a unique opportunity to understand the cues that regulate neonatal behaviour in a mammal. The female nurtures her litter for 4–5 min once a day⁷ during the 2 weeks after birth: this is therefore less than 0.35% of her time. Because the survival of the pups is conditional on milk intake on the first 2 days of life⁸, the pups need a reliable sensory tether for the rapid location of nipples, and competent behaviour to obtain milk successfully in a context of harsh competition between littermates. Rabbit pups are endowed with keen chemosensory and tactile abilities linked with a typical head-searching pattern that ends in the grasping of a nipple within 3–5 s (refs 9, 10). The release of this behaviour is under the main control of chemical cues emitted by lactating females on the nipples^{10,11} and in milk¹². As these skin and milk signals are functionally equivalent¹³, we focused our analytical effort on rabbit milk.

The volatiles in milk were extracted, trapped and then desorbed into a gas chromatograph (GC) equipped with a sniffing device,

permitting concurrent detection by neonatal rabbits and the flame ionization detector (FID). The pups subjected to such gas chromatography–olfaction (GCO) tests responded either by short-range searching motions of the head directed to the sniff-port, or by attempts to seize it orally (Fig. 1a). The cumulative frequency of both responses allowed us to determine behaviourally active regions on the chromatographic profiles (Fig. 1b). Although some compounds could not be identified because of uninterpretable mass spectra caused by concentration or co-elution problems, most of the GC peaks corresponding to the behavioural peaks were identified by GC–mass spectrometry (GC–MS). The 21 compounds identified were screened for behavioural activity with a bioassay presenting them (aqueous dilution of 1 µg ml^{−1}) on a glass rod (Fig. 2a). One volatile, 2-methylbut-2-enal (2MB2; Fig. 1c), released both responses to a considerably greater extent than any of the other volatiles ($\chi^2 > 50$, $P < 0.001$, for all comparisons) and than the solvent alone ($\chi^2 > 60$, $P < 0.001$) (Fig. 2b).

The pup responsiveness to 2MB2 was dependent on concentration. Its intensity-tuning curve established with the glass-rod test revealed optimal effectiveness between dilution steps 10 ng ml^{−1} and 1 µg ml^{−1} (Fig. 3a). Thus, all subsequent behavioural assays used a constant concentration of 2MB2 (1 µg ml^{−1}, in water). The 20 remaining compounds isolated from rabbit milk were behaviourally inert at all concentrations between 100 pg ml^{−1} and 10 mg ml^{−1} (G.C., D.L., G.P. and B.S., unpublished observations).

Additional experiments were conducted to ascertain further that 2MB2 is a key compound in rabbit milk for pups. First, all detectable impurities co-occurring with commercial 2MB2 (namely acetic acid, benzoic acid, 3-methylbut-2-enoic acid and nonanal) were assayed on glass rods and excluded because they had no or negligible effectiveness. Second, when such 2MB2 was assayed by GCO with a polar capillary column, the timing of pup responses corresponded with the retention time of natural 2MB2. Third, the differential separative properties of an apolar GC column were used to check for compounds eluting with 2MB2; in that case too the timing of pup responsiveness matched the retention time of natural 2MB2.

To corroborate the effect of 2MB2, we related its concentration in the milk headspace with the magnitude of the behavioural effect of milk. After standing for 60 min at ambient temperature, rabbit milk loses its effectiveness to evoke pup responsiveness^{12,14}. Thus, if 2MB2 is the key compound in milk, the activity loss of the latter would predict a decrease in 2MB2 concentration. The 2MB2 concentration in fresh milk decreased on standing for 30 and 90 min (analysis of variance comparing 0, 30 and 90 min, $F_{2,13} = 8.2$, $P < 0.01$; 0 and 30 min, $F_{1,11} = 7.1$, $P < 0.05$; 30 and 90 min, $F_{1,6} = 4.7$, $P < 0.07$; 0 and 90 min, $F_{1,9} = 36$, $P < 0.01$) and was clearly associated with a drop of neonatal searching–grasping responsiveness ($\chi^2 > 5$, $P < 0.05$ for all comparisons; Fig. 3c). Furthermore, when milk was rendered inactive by headspace extraction (response frequency to milk before and after deodorization, $\chi^2 > 19$, $P < 0.001$ for both responses), its behavioural activity was reinstated by the addition of 2MB2 (response frequency to deodorized versus 2MB2-enriched milks, $\chi^2 > 26$, $P < 0.001$ for both responses; to fresh versus 2MB2-enriched milks, $\chi^2 = 4.8$, $P < 0.05$, for searching; Fig. 3b).

A series of experiments was then undertaken to verify step by step whether 2MB2 carries the psychobiological properties that define a pheromone. Although there is a long-standing semantic debate about the definition of a pheromone in mammals, we used the most restrictive meaning of the concept^{15,16}. Thus, five operational criteria were considered to assess whether 2MB2 can qualify as a pheromone: chemical simplicity of the signal; unambiguous, morphologically invariant, and functionally obvious behavioural response of the receiver; high selectivity of stimulus–response coupling; species specificity of reception; and unconditional stimulus–response coupling. A sixth criterion has been added by us that relates to the species specificity of emission of the odour cue.

A dilute aqueous solution of 2MB2 triggered a clearly defined, stereotyped response in newborn pups that was not immediately differentiable from that elicited by fresh rabbit milk. Thus, a single compound from milk releases the same behavioural effect as whole milk, which carries more than 150 volatiles. Further, this response was not influenced by the mode and general context of presentation, as 2MB2 released stereotyped searching–grasping responses in the GCO assay (Fig. 1a), in the assay with the glass rod (Fig. 2a) and in a two-choice arena assay (G.C., D.L., G.P. and B.S., unpublished observations).

The selective bioactivity of 2MB2 was tested against 20 odorants absent from rabbit milk that were either chosen arbitrarily, previously identified in rabbit skin-gland secretions¹⁷ or known to be active in other mammalian newborns^{18,19}. The behavioural effectiveness of these stimuli was either null or significantly lower than that of 2MB2 (response frequencies to 2MB2 versus other compounds, $\chi^2 > 60$, $P < 0.001$, for both responses and in all cases; Fig. 4a). Thus, 2MB2 activity cannot be ascribed to non-specific arousal or novelty effects that would be elicited by any odorant. The selectivity of 2MB2 was further assessed against odorants that pups had previously encountered in highly reinforcing conditions, namely dietary aromas transferred to the fetal

environment and milk²⁰. Same-breed pups (New Zealand $\times$ Californian) assayed in five different colonies with distinct diets indicated that the bioactivity of 2MB2 did not depend on the individual odour type of the females or on their diet. Furthermore, when pregnant and lactating females of a same colony were fed with two exclusive diets, their pups responded at the same rate to 2MB2 (G.C., D.L., G.P., L. Fortun-Lamothe, F. Lebas and B.S., unpublished observations), reinforcing the notion that the efficacy of 2MB2 is independent of the individual bouquet of volatiles experienced in the amnion or milk.

The generality of 2MB2 activity was assessed in pups from other rabbit strains and breeds including angora, castor, chinchilla, laghmore, butterfly and Belgian hare. 2MB2 assays elicited searching–grasping in more than 80% of those pups (Fig. 4b), showing that genotype has no effect on its releasing power. 2MB2 reception specificity was also examined in newborns belonging to taxa having more or less phylogenetic distance from *Oryctolagus*. Newborns of a closely related lagomorph, *Lepus europaeus*, and of murine rodents (*Rattus rattus*, *Mus musculus*) and a carnivore (*Felis catus*) were tested: the outcome was negative, making clear the species specificity of 2MB2 (Fig. 4b).

The above demonstration of species-level generality and speci-

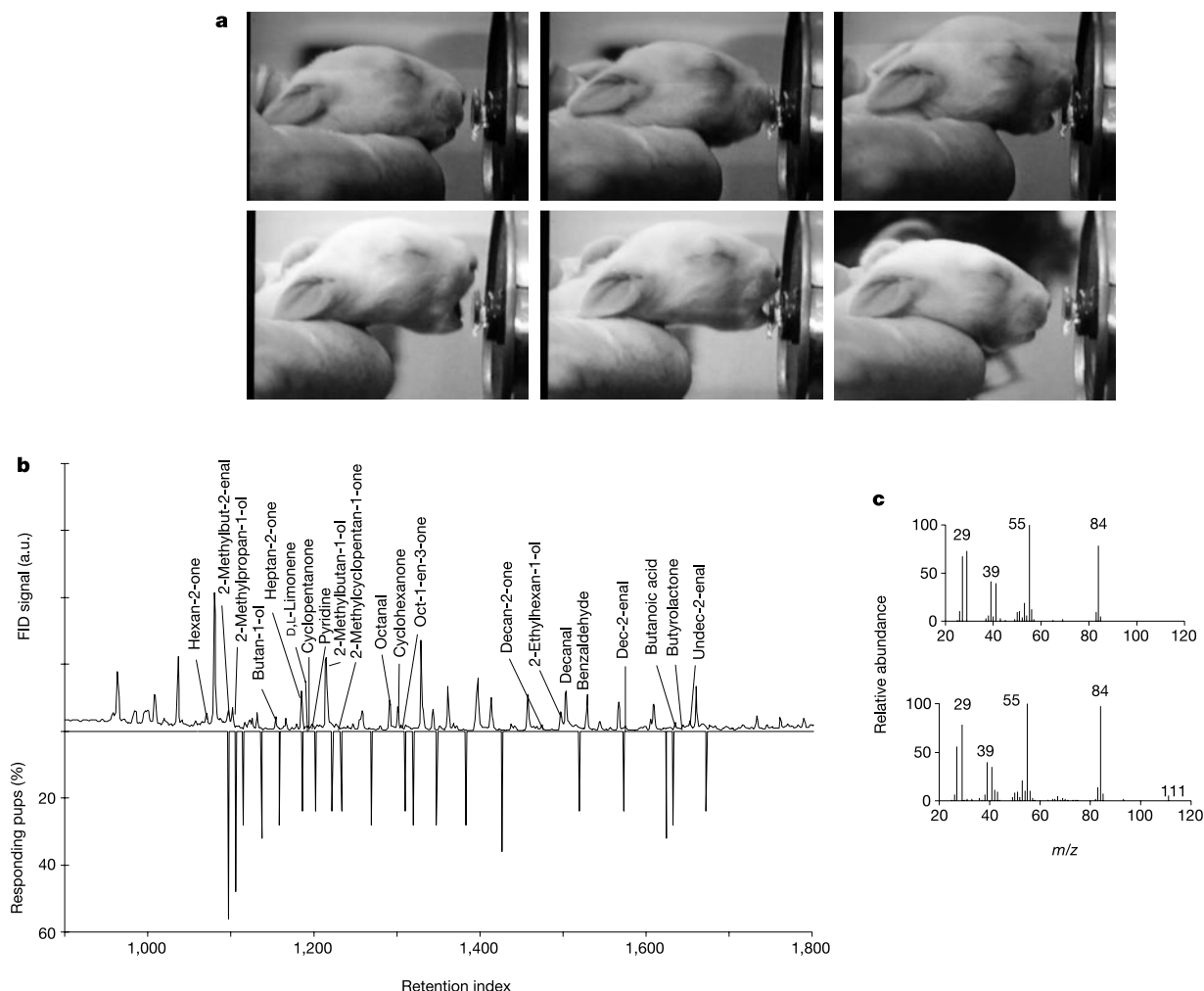


Figure 1 The GCO assay. **a**, Sequence (duration 5 s) of a 2-day-old pup's searching–grasping response directed to the glass funnel of the GC sniff-port. **b**, Typical chromatogram of rabbit milk effluvium (upper panel) and concurrent percentage of pups responding with searching–grasping responses (lower panel; inverted scale). The regions

of the chromatogram eliciting more than 20% of responses (summed across 25 GCO runs) and the compounds eluting in these regions are shown. **c**, Mass spectra of 2MB2 of commercial origin (upper panel) and from rabbit milk (lower panel). Both scans were made at a retention index value of 1,096.

ficity of 2MB2 perception in *Oryctolagus* newborns calls for the assessment of the generality of 2MB2 secretion and release in *Oryctolagus* females, irrespective of their breed and particular food intakes. The GC profiles of milk volatiles from different individual females indicated that 2MB2 is invariably present among a background of more variable volatiles. 2MB2 might thus be generated by species-specific biosynthetic processes. We further examined whether 2MB2 was externalized in rabbit milk only or whether it was ubiquitous in mammalian milk. We therefore assayed rabbit newborns with glass rods dipped in freshly obtained colostrum and/or milk from rats, sheep, cows, horses and humans. In no case did any of these milks elicit the typical responses. This was confirmed with the easily accessible cow milk, in the headspace of which no 2MB2 was detected by GC–MS.

Finally, we considered how far the development of the neonatal response to 2MB2 depended on learning processes. First, the activity of 2MB2 was assessed in pups that had never been exposed to it postnatally. Vaginally delivered newborns deprived of contact with the doe's belly or milk, and caesarean-delivered pups (1 day preterm) were assayed with 2MB2 on a glass rod 10–15 min after birth. These naive newborns displayed searching–grasping responses at the same rate as pups exposed to the doe (Fig. 4c). Thus, to be fully effective, 2MB2 does not require natal or postnatal exposure or reinforcement. Second, because the specific 2MB2–response coupling is established by at least 1 day before birth, it might be derived from prenatal exposure to 2MB2. We assessed this possibility in assaying the substrates that directly contact the developing chemoreceptors. When presented with glass rods bearing blood of either pregnant females (last day of gestation) or lactating females (day 3 postpartum), or amniotic fluid (collected on gestational days 12 or 28), pups remained inactive. Further, no 2MB2 was detected in these substrates by GC–MS. Therefore the

development of 2MB2 activity apparently does not depend on prenatal induction through direct exposure to the stimulus.

2MB2 is thus a volatile compound from rabbit milk that qualifies as a pheromone according to the five criteria mentioned above. It is a single compound that selectively triggers a stereotyped behavioural response of clear functional significance in newborns and of obvious benefit to both receiver and sender. The neonatal coupling between 2MB2 reception and the searching–grasping response is species specific and it does not seem to derive from either postnatal or prenatal acquisition. In addition, 2MB2 is present in milk irrespective of the female's diet, suggesting that it is produced *de novo* in the mammary tract. We therefore name it mammary pheromone. It is noteworthy that this single compound accounts fully for the activity of the complex mixture of volatiles carried in milk. This situation contrasts with the great majority of the putative pheromones known in mammals, which are most often composed of mixtures that lose their releasing power after fractionation^{15,16,21}.

Investigations on other mammals have characterized odorants associated with the mammary gland or nipple and it has often been claimed that these odorants have pheromonal properties. However, these studies either were restricted exclusively to behaviour testing without supporting chemical analyses²², or identified putative pheromonal compounds without a full verification of the criteria that would justify their unambiguous designation as pheromones²³. In addition, in the rare cases when natural odorants eliciting nipple attachment were shown, they clearly did not originate in the mammary tract itself but were spread there from other body regions (for example saliva or amniotic fluid) through the female's and/or the newborns' behaviour²³. Thus, the present work definitively

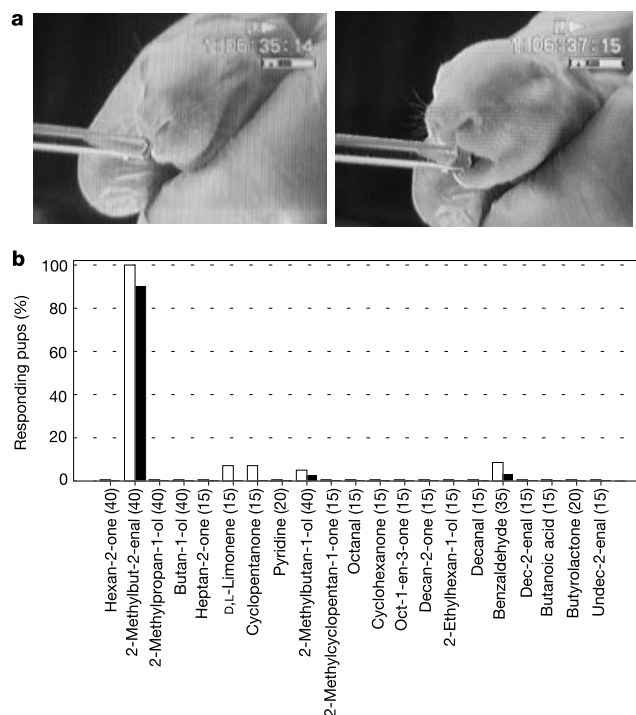


Figure 2 Screening, by the glass-rod assay, of milk volatiles presumed to have behavioural activity. **a**, A 2-day-old pup at rest (left) and exhibiting grasping (right) to a glass rod carrying 2MB2. **b**, Frequency of searching (open bars) and seizing (solid bars) directed to the glass rod carrying one of the 21 compounds identified in milk. Numbers in parentheses indicate the numbers of pups tested.

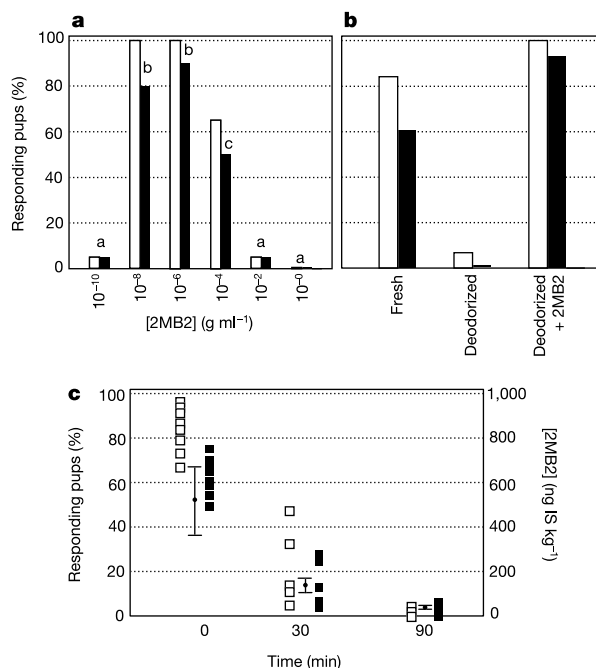


Figure 3 The key impact of 2MB2 in milk (glass-rod assays; open bars and squares, searching; solid bars and squares, grasping). **a**, Relation between 2MB2 concentration (g ml^{-1}) and percentage of pups responding. For each response, distinct letters indicate statistical differences between concentrations ($P < 0.05$). Twenty pups were used for each dilution step. **b**, Frequency of pups responding to the odours of fresh milk, deodorized milk and milk spiked with 2MB2. Fifteen pups were used for each test. **c**, Frequency of pups (squares) responding to the odour of milks standing for various delays after milking (mean duration of milking: $15.6 \pm 5 \text{ min}$ ($\pm \text{s.d.}$)) and concentration (points and error bars; means $\pm \text{s.d.}$) of 2MB2 in the headspace of these samples. All milks were sampled from two females. Fifteen pups were used for each test.

reports the existence of a mammary pheromone in a mammalian species.

This finding has several broader implications for future research. First, such key compounds releasing clear behavioural responses are pressing needed to unravel further the general principles of olfactory functioning, plasticity and development in vertebrate models. At the systems level, the pheromone-behaviour coupling will permit the explanation of the chemosensory subsystem mediating its detection. It should then help in deciphering the specificity of the activity pattern elicited in these chemosensory pathways from peripheral to higher-order neural levels. At the cellular level, this ligand should permit avoiding the use of more complex natural stimuli (see ref. 24, for example) and might be applied as a tool for the understanding of the molecular processes involved in its capture and transduction. This approach should allow an assessment of whether certain nasal chemoreceptor cells are a part of domain-specific modules dedicated to the processing of a single compound (see ref. 25, for example). Finally, this key compound will allow the systematic exploration of structure-activity relationships at organismic, systems and cellular levels^{26,27}.

Second, the above finding spurs the scrutiny of whether this method of mother-offspring information exchange emerged with

particular life-history patterns (such as the 'absentee' parenting of the European rabbit) or whether it is a general strategy linked with lactation in mammals. The behavioural-chemical methodology developed here could be used to screen for specialized chemocommunicative means associated with mammaries in comparative and phylogenetic perspectives. □

Methods

Animals

The animals under study originated from five colonies (mainly ENESAD, Dijon; Cunifrance, Epeigné-sur-Dême; Inra, Nouzilly, Le Magneraud and Toulouse) and were kept under standard conditions^{8,11,14}. Does and their litters were submitted to a schedule mimicking nursing once a day⁸. All experiments followed the rules of French Ministries (Agriculture; Research and Technology). Vaginally delivered pups were tested on days 0–4 after birth and caesarean-delivered pups were tested on day 30 after conception. Pups were assayed 22–24 h after last nursing unless stated otherwise.

Extraction of volatiles and GCO assay

A nitrogen stream was flowed (100 ml min⁻¹ for 2 h) through fresh milk (held at 30 °C) and directed to a Tenax trap (200 mg type TA, 20–35 mesh), which was then fitted to a TCT-CP4010 injector (Chrompack) placed on a 5890 GC (Hewlett Packard). The trapped volatiles were heat-desorbed (240 °C for 20 min) and subsequently condensed into a liquid-nitrogen-cooled (–130 °C) section of capillary tubing preceding a DB wax column (30 m length, 0.32 mm inside diameter, 0.5 µm phase thickness; J&W Scientific). This section of capillary tubing was then heated rapidly to 250 °C and the GC was run in temperature-programmed mode (40–240 °C at 5 °C min⁻¹; carrier gas H₂, velocity 50 cm s⁻¹).

At the end of the capillary column, the stream of volatiles was split into two equal flows directed by fused silica capillaries to the FID and to a sniff-port²⁸ ending in a glass funnel. The capillary (0.32 mm internal diameter) on the GC column ended at the entry of the glass funnel, whose temperature was 50 °C; at a distance of 5 mm from the funnel the temperature was 23 °C. The sniff-port allowed olfactory sampling by the pups, whose muzzle was positioned 5 mm from the funnel by trained experimenters wearing gloves (Fig. 1a). For each 35-min long GCO run, 35 pups were assayed (from 7 litters, 5 pups per litter, excluding the litter of the milk donor female). Pups were alternated every 1 min with minimal interruption periods (less than 2 s) between consecutive pups; after the first GCO run, a plus or minus 30-s lag was entered at the onset of each GCO run so that the interruptions were not coincident with the same time periods in all GCO runs (that is, 18 runs made at *t* 0 s and 17 runs at *t* –30 s). Three experimenters ran the procedure: one ran the GC and recorded pup responsiveness on both computer and GC paper traces; two others, who were blind to the progress of the GC, alternatively exposed the pups to the sniff-port and reported the pups' responses to the GC operator.

We ran 25 GCO assays on the headspace of milk from 25 females. Among the 875 pups (from 78 litters) used, 656 were tested once and 219 were tested twice (with a minimal inter-assay delay of 5 h).

To verify the activity of 2MB2, additional GCO runs were undertaken by injecting commercial 2MB2 (1 µl, 25 µg ml⁻¹ in dichloromethane) into a GC equipped with either a polar DB wax column or an apolar DB5 column (30 m, 0.32 mm, 1 µm; J&W Scientific). Both series of GCO were run using these conditions.

Identification of volatiles by mass spectrometry

A Hewlett Packard 5890 GC equipped with a TCT-CP4010 injector (same conditions as above) was used with a R10-10C MS (Nermag; ionization energy 70 eV; scan frequency range *m/z* 25–300 per 800 ms). The candidate compounds were identified from the Wiley library of mass spectra and were purchased from Aldrich.

Assay with glass rod

The assay consisted of a 10-s presentation of a glass rod (20 cm long, 0.4 cm in diameter) carrying the substance to be tested^{12,14,20}. The glass rod was positioned 5 mm in front of the pup's nares by trained experimenters wearing gloves. The variables were the frequency of pups responding by head-searching movements and/or oral grasping of the rod (Fig. 2a).

Dosage of active compound, and milk deodorization

Different milk samples (30 g) were poured into a 250-ml flask immersed in a 30 °C water bath; 3-methylbutanoic propyl ester (24 ng) was added as an internal standard. A nitrogen stream was flowed (50 ml min⁻¹, 1 h) above the milk surface and directed to a Tenax trap used as above. The GC was coupled with the MS used in single-ion monitoring mode. For quantification, two fragments were selected for 2MB2 (*m/z* 55–84) and for the internal standard (*m/z* 71–89). A calibration curve was established by measuring the MS response to increasing concentrations of commercial 2MB2 and the constant concentration of the internal standard in milk. The concentration of 2MB2 in milk was extrapolated from this calibration curve.

Deodorization of fresh milk was performed by bubbling nitrogen through it (100 ml min⁻¹ for 2 h). Thereafter the milk was spiked with 2MB2 (final concentration 1 µg ml⁻¹).

Collection of biological samples

The rabbits (0–4 d post partum) were milked with Lebas' or Schley's milking device^{29,30}. Blood and amniotic fluid were obtained during studies on rabbit reproductive physiology conducted at Inra, Toulouse. Immediately after the animal's death, venous blood and

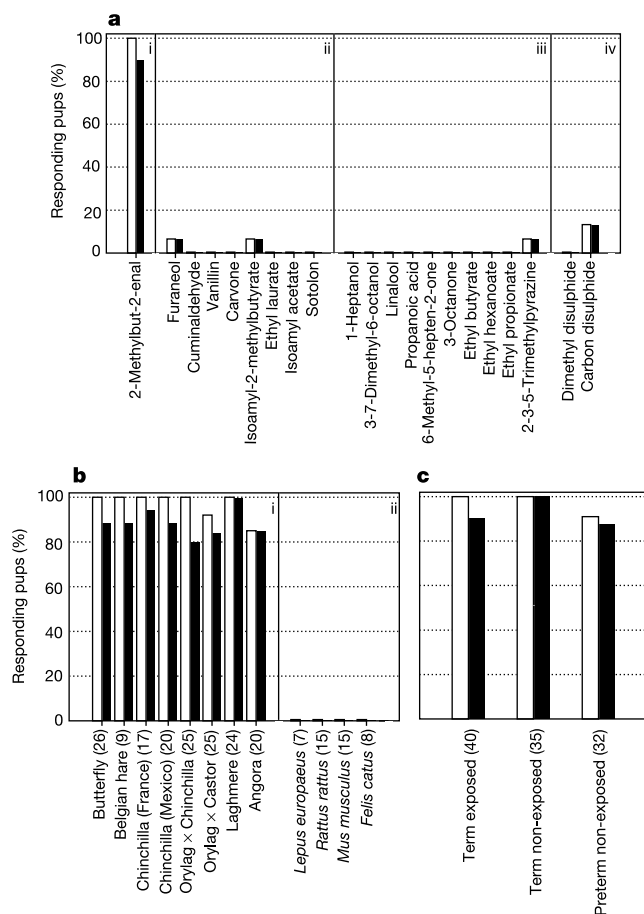


Figure 4 Selectivity, specificity and independence from learning of pup responsiveness to 2MB2 (glass-rod assays; open bars, searching; solid bars, grasping). **a**, Selective bioactivity of (panels from left to right) 2MB2, odorants chosen arbitrarily, odorants excreted by adult rabbits, and odorants known to be active in rat pups. Forty pups were used for each test for 2MB2 and 15 for each test for all other compounds. **b**, Frequency of newborns responding to 2MB2 in different rabbit breeds (left) and lagomorph or non-lagomorph mammals (right). **c**, Frequency of pups responding to 2MB2 as a function of birth timing (term, preterm) and exposure to maternal abdomen and milk (exposed, non-exposed) before the assay. In **b** and **c**, numbers in parentheses indicate the number of pups in each group. All artificial odorants were diluted to 1 µg ml⁻¹ in water.

amniotic fluid were sampled and frozen at once (-80°C). Lactal secretions were milked from non-rabbit species (colostrum from two rats (1 d post partum); colostrum from two ewes (15 h post partum) and milk from two other ewes (8 d post partum); milk from three cows (late gestation); colostrum and milk from six mares (10–15 h and 6 d post partum); milk from one woman (21 d post partum)) and tested on groups of 20 rabbit pups (from four litters, five per litter, aged 2–3 d).

Received 18 November 2002; accepted 31 March 2003; doi:10.1038/nature01739.

- Clutton-Brock, T. H. *The Evolution of Parental Care* (Princeton Univ. Press, 1991).
- Rosenblatt, J. S. & Snowdon, T. C. *Parental Care, Evolution, Mechanisms and Adaptive Significance* (Academic, Orlando, 1996).
- Rosenblatt, J. S. Olfaction mediates developmental transition in the altricial newborn of selected species of mammals. *Dev. Psychobiol.* **16**, 347–375 (1983).
- Leon, M. & Moltz, H. Maternal pheromone: Discrimination by pre-weanling albino rats. *Physiol. Behav.* **7**, 265–267 (1971).
- Blass, E. M. & Teicher, M. H. Suckling. *Science* **210**, 15–22 (1980).
- Macfarlane, A. J. Olfaction in the development of social preferences in the human neonate. *Ciba Found. Symp.* **33**, 103–117 (1975).
- Zarrow, M. X., Denenberg, V. H. & Anderson, C. O. Rabbit: Frequency of suckling in the pup. *Science* **150**, 1835–1836 (1965).
- Coureaud, G. et al. Immediate postnatal sucking in the rabbit: Its influence on pup survival and growth. *Reprod. Nutr. Dev.* **40**, 19–32 (2000).
- Schley, P. Geruchssinn und Saugverhalten bei Jungkaninchen. *Kleintier Praxis* **26**, 261–263 (1981).
- Hudson, R. & Distel, H. Nipple location by newborn rabbits: Evidence for pheromonal guidance. *Behaviour* **85**, 260–275 (1983).
- Coureaud, G. & Schaal, B. Attraction of newborn rabbits to abdominal odors of adult conspecifics differing in sex and physiological state. *Dev. Psychobiol.* **36**, 271–281 (2000).
- Keil, W., von Stralendorff, F. & Hudson, R. A behavioral bioassay for analysis of rabbit nipple-search pheromone. *Physiol. Behav.* **47**, 525–529 (1990).
- Coureaud, G., Schaal, B., Langlois, D. & Perrier, G. Orientation response of newborn rabbits to odours of lactating females: Relative effectiveness of surface and milk cues. *Anim. Behav.* **61**, 153–162 (2001).
- Coureaud, G. *Régulation Olfactive de la Prise Lactée chez le Lapereau: Caractérisation Éthologique et Chimique d'un Signal Pheromonal*. Thesis, Univ. Paris 13 (2001).
- Beauchamp, G. K., Doty, R. L., Moulton, D. G. & Mugford, R. A. in *Olfaction, Behavior, and Mammalian Reproduction* (ed. Doty, R. L.) 143–160 (Academic, New York, 1976).
- Johnston, R. E. in *The Neurobiology of Taste and Smell*, 2nd edn (eds Finger, T. E., Silver, W. L. & Restrepo, D.) 101–127 (Wiley-Liss, New York, 2000).
- Goodrich, B. S., Hesterman, E. R., Shaw, K. S. & Mykityowycz, R. Identification of some volatile compounds in the odor of fecal pellets of the rabbit, *Oryctolagus cuniculus*. *J. Chem. Ecol.* **7**, 817–827 (1981).
- Pedersen, P. A. & Blass, E. M. in *The Development of Perception: Psychobiological Perspectives* (eds Aslin, R. N., Alberts, J. R. & Petersen, M. R.) 359–381 (Academic, New York, 1981).
- Galef, B. G., Mason, J. F., Preti, G. & Bean, N. J. Carbon disulfide: A semiochemical mediating socially-induced diet choice in rats. *Physiol. Behav.* **42**, 119–124 (1988).
- Coureaud, G., Schaal, B., Hudson, R., Orgeur, P. & Coudert, P. Transnatal olfactory continuity in the rabbit: Behavioral evidence and short-term consequence of its disruption. *Dev. Psychobiol.* **40**, 372–390 (2002).
- Muller-Schwarze, D. et al. in *Chemical Signals in Vertebrates 4* (eds Duvall, D., Muller-Schwarze, D. & Silverstein, R. M.) 561–570 (Plenum, New York, 1986).
- Morrow-Tesch, J. & McGlone, J. J. Sensory systems and nipple attachment behavior in neonatal pigs. *Physiol. Behav.* **47**, 1–4 (1989).
- Teicher, M. H. & Blass, E. M. First suckling response of the newborn albino rat: The roles of olfaction and amniotic fluid. *Science* **198**, 635–636 (1977).
- Sam, M. et al. Odorants may arouse instinctive behaviours. *Nature* **412**, 142 (2001).
- Zou, Z., Horowitz, L. F., Montmayeur, J. P., Snapper, S. & Buck, L. B. Genetic tracing reveals a stereotyped sensory map in the olfactory cortex. *Nature* **414**, 173–179 (2001).
- Johnson, B. A. & Leon, M. Odorant molecular length: One aspect of the olfactory code. *J. Comp. Neurol.* **426**, 330–338 (2000).
- Uchida, N., Takahashi, Y., Tanifuji, M. & Mori, K. Odor maps in the mammalian olfactory bulb: Domain organization and odorant structural features. *Nature Neurosci.* **3**, 1035–1043 (2000).
- Abbott, N., Etievant, P., Issanchou, S. & Langlois, D. Critical evaluation of two commonly used techniques for the treatment of data from extract dilution sniffing analysis. *J. Agric. Food Chem.* **41**, 1698–1703 (1993).
- Schley, P. *Untersuchung zur kuenstlichen Aufzucht von Hauskaninchen*. Thesis, Univ. Giessen (1976).
- Lebas, F. Description d'une machine à traire les lapines. *Ann. Zootech.* **19**, 223–228 (1970).

Acknowledgements We thank J. L. Le Quéré for support and advice; P. Coudert, J. Ponceau, J. L. Vrillon and P. Mercier for giving access to their colonies; P. Orgeur, J. P. Drouet, M. Jouanno, F. Lebas, L. Fortun-Lamothe, M. Theau-Clément, G. Thébaud, G. Bolet, A. Locatelli, Y. Breuzin and J. P. Francinot for contributions; D. Licois and R. Hudson for providing Lebas' and Schley's milking devices; R. L. Doty, V. H. Denenberg, A. Holley, J. P. Montmayeur, P. Coudert, R. Hudson, P. Salin and H. Distel for discussions; and R. Langlois, E. Hertling and C. Ferdenzi for logistic support. This work was conducted in part while B.S. and G.C. were at Unité de Physiologie de la Reproduction, CNRS-Inra, Nouzilly, France. G.C. was sponsored by Région Poitou-Charente-Inra and the Fyssen Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.S. (schaal@cesg.cnrs.fr).

Spontaneous muscle twitches during sleep guide spinal self-organization

Per Petersson, Alexandra Waldenström, Christer Fåhræus & Jens Schouenborg

Section for Neurophysiology, Department of Physiological Sciences, BMC F10, Lund University, S-221 84 Lund, Sweden

During development, information about the three-dimensional shape and mechanical properties of the body is laid down in the synaptic connectivity of sensorimotor systems through unknown adaptive mechanisms. In spinal reflex systems, this enables the fast transformation of complex sensory information into adequate correction of movements. Here we use a computer simulation to show that an unsupervised correlation-based learning mechanism, using spontaneous muscle twitches, can account for the functional adaptation of the withdrawal reflex system. We also show that tactile feedback resulting from spontaneous muscle twitches during sleep^{1–3} does indeed modify sensorimotor transformation in young rats in a predictable manner. The results indicate that these twitches, corresponding to human fetal movements⁴, are important in spinal self-organization.

The immense computational complexity of generating, coordinating and controlling compound multi-joint movements in mammals has led to the assumption that even relatively simple sensorimotor tasks such as the nociceptive withdrawal reflex (NWR) could be processed only in widely distributed multilayer

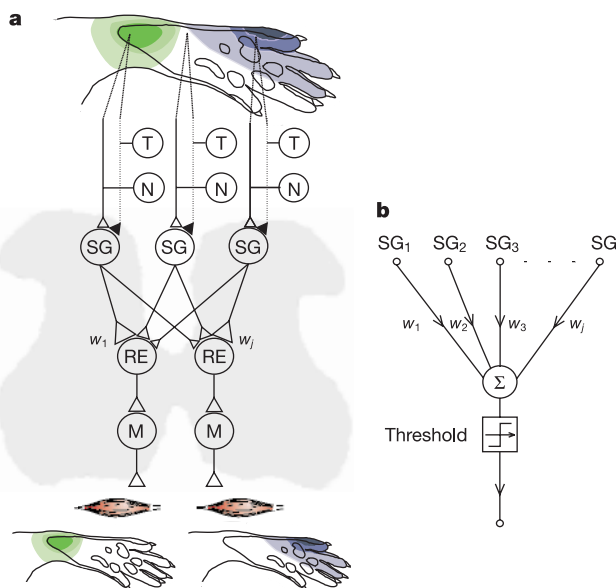


Figure 1 Proposed organization of a reflex module. **a**, The features are as follows. First, many primary afferents feed into the reflex circuit monosynaptically or oligosynaptically (continuous or dotted lines); second, tactile (T) and nociceptive (N) inputs converge in substantia gelatinosa (SG) before reaching the reflex-encoding neurons (REs) located in the deep dorsal horn; third, SG to RE connections are specifically weighted (w_j = synaptic weights); fourth, REs have strong connections to motor neurons activating the muscle(s). The receptive and withdrawal fields of the gastrocnemius (green) and peroneus longus (blue) muscles, respectively, are shown. Filled boutons, inhibitory synapses; open boutons, excitatory synapses. **b**, A Rosenblatt perceptron for comparison.

networks⁵. However, recent data⁶ instead indicate that the spinal withdrawal reflex system has a modular organization, in which each withdrawal reflex module acts primarily on one muscle (Fig. 1a). For hindlimb reflexes, the sensory input from the skin to a module has a weight distribution that is an imprint of the withdrawal efficiency of the muscle when the limb is in a standing-like position, with the foot in contact with the ground⁶. This imprint is laid down on the reflex pathways through extensive postnatal adjustments; erroneous connections are either eliminated or decreased, and the strength of adequate connections becomes proportional to withdrawal efficiency⁷. This process takes approximately a week in the rat and occurs during the first three postnatal weeks depending on the body part^{7,8} (Fig. 2). The sensorimotor transformation can adapt to neonatally induced alterations of either peripheral innervation⁹ or movement patterns¹⁰.

Here we propose that each module is self-organizing and learns about its withdrawal efficiency by probing the tactile feedback that arises from spontaneous activity in its reflex interneurons, activating the associated principal muscle. The temporal correlation between the initial activity in the reflex interneurons triggering the movement and the ensuing tactile feedback would allow a reversed type of hebbian learning (with the postsynaptic activity in reflex interneurons preceding the afferent input). This unsupervised learning principle is referred to below as motor-directed somatosensory imprinting (MDSI) (Fig. 3b). The bases for this hypothesis are the following observations. First, the developmental adaptation is dependent on tactile input, but independent of nociceptive input⁸. Second, during early development the movement repertoire is dominated by spontaneous muscle twitches³ that occur during sleep^{1,2,11}. The spontaneous twitches, characterized by brief contractions in atonic muscles¹, usually result in increased or decreased skin contact with the environment, causing altered tactile input. Many of these twitches are dominated by a single muscle^{11–13} (see also Supplementary Information). Third, the generators of the spontaneous activity are, at least partly, intrinsic to the spinal cord because spontaneous twitches are unaltered by deafferentation¹⁴ and are not abolished by spinal transection³. Last, the rat brain and its connections with the spinal cord¹⁵ are immature during the first few postnatal weeks. A direct instructive role of supraspinal systems in spinal motor learning is therefore unlikely. However, a permissive role is conceivable because spinal transection at birth interferes with NWR adaptation¹⁶.

To test the plausibility of this hypothesis, a computer simulation of MDSI was performed. The simulation was based on published data on plantar hindpaw excitatory receptive fields of six NWR modules in the rat, and skin displacement patterns due to single hindlimb muscle contractions obtained in a standing-like position⁶

(similar to the normal position of the hindlimbs during twitching; see Supplementary Information). The network organization consisted of an input layer of interneurons in substantia gelatinosa that receive nociceptive and tactile input from the same skin area¹⁷, and an output layer of deep dorsal-horn neurons encoding the strength of the withdrawal reflex, termed reflex encoders, that project to the motor neurons (Fig. 1a). This network architecture is similar to a classical Rosenblatt perceptron¹⁸ (Fig. 1b). Oja's rule¹⁹ (see Methods) was chosen to describe adjustments of the connection weights between substantia gelatinosa and the reflex encoders. This learning rule differs from classical hebbian learning in that it includes a synaptic scaling factor that readjusts synaptic weights gradually when uncorrelated postsynaptic activity is present and is therefore self-normalizing²⁰. Note that in MDSI the relative amplitude, rather than the absolute amplitude, of the tactile feedback determines the change in connection strengths such that connections from skin areas from which stronger tactile input is received after a twitch will weaken compared with those from skin areas with less tactile input. Correlations between synaptic weight maps of the developing receptive fields and the movement pattern of each muscle were calculated at different time points during the adaptation process (Fig. 3a). The learning curves for the six modules that were obtained from the simulation (Fig. 3c) were similar to previous experimental data (Fig. 2b). The model was relatively insensitive to the choice of parameter values (such as twitching probability, learning rate, sensory noise level and initial synaptic weights; see Methods). For the data shown in Fig. 3, inhibitory feedback connections of tactile input to the substantia gelatinosa cells (as shown in Fig. 1) were modelled¹⁷. However, if tactile input instead exerts excitatory effects on the substantia gelatinosa cells²¹ then anti-Oja learning (replacing η with $-\eta$ in Oja's rule) yields similar relative adult weight distributions and comparable learning curves. By contrast, unsupervised feedforward learning (often assumed to underlie the fine tuning of sensory representations²²) in the same six perceptron-like networks as used in MDSI, based on the same input

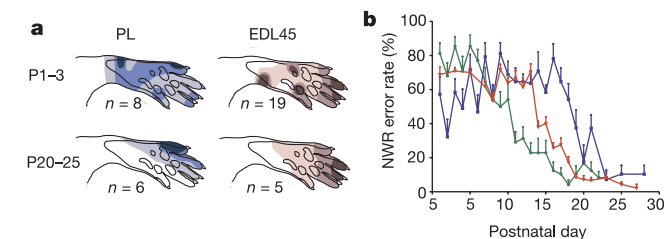


Figure 2 Developmental adaptation of NWRs. **a**, Averaged receptive fields of the PL and EDL45 muscles obtained at different postnatal ages by using CO₂-laser stimulation (intensity twofold the reflex threshold; *n* is the number of rats; see Methods for abbreviations). **b**, Reflex adaptation curves. Graph showing error rate of tail NWRs at different postnatal ages. The tail was stimulated twice, distally, on both lateral sides during P1–P28 in two litters (blue, *n* = 8; green, *n* = 12) and six times on each side in the third litter (red, *n* = 13). (Modified from refs 8, 9.)

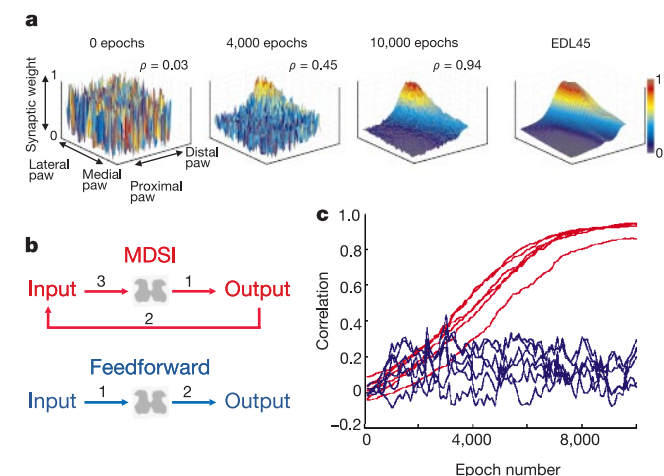


Figure 3 Simulated developmental adaptation of receptive fields for six muscles. **a**, Simulated synaptic weight distributions of a receptive field (EDL45) at different time points during reflex adaptation. **b**, The two simulated learning principles: MDSI and feedforward learning. The sequence of events is numbered. **c**, Simulated reflex adaptation curves. Correlations between synaptic weight maps and movement patterns (597 skin sites) at different time points for the respective muscles with the use of Oja's learning rule for six networks during MDSI (red) or feedforward learning (blue). Simulated (MDSI) and experimental adult correlations, respectively: $r_{EDL23} = 0.94$ and 0.91 , $r_{EDL45} = 0.93$ and 0.95 , $r_G = 0.85$ and 0.95 , $r_{PB} = 0.94$ and 0.88 , $r_{PL} = 0.93$ and 0.91 , $r_{TA} = 0.93$ and 0.90 . (Parameter values and abbreviations are given in Methods.)

patterns and Oja's rule, did not result in correlations stronger than 0.4 (Fig. 3c).

The agreement between the simulation and experiments indicated that our model had high validity. Nevertheless, the role of spontaneous muscle twitches during sleep for spinal self-organization had to be evaluated experimentally. This was done in behavioural experiments in which artificial sensory feedback was given shortly after spontaneous muscle twitches. For technical reasons we studied tail NWRs in these experiments. An automatic imaging system that triggered air-puff stimulation was developed. This stimulation was directed to either side of the tail upon detection of isolated, rapid tail movements. Before training began, the animals were randomized to either air-puff stimulation from the side of the set-up that the tail was moving towards or from the side it was moving away from. Because an adequate NWR can cause an increased input from the skin area moving towards external objects but should not cause an increased input from the skin area withdrawn, these two air-puff stimulations are referred to as 'normal' and 'aberrant', respectively (Fig. 4a). Rats were trained in the system for 2 h a day (mean $\pm$ s.d., 944 ± 526 stimulations) during the period when adequate tail NWRs are normally learned (postnatal ages 12–17 days (P12–P17))⁸. The effect of training was examined by comparing the error rate of NWR responses before (less than 1 h) and after (less than 10 min) each training session. Reflexes were elicited by cutaneous nociceptive CO₂-laser stimulation of either side of the tail in the horizontal plane. Reflex movements towards the laser source were classified as erroneous. After aberrant air-puff conditioning (895 ± 493 stimulations), the mean NWR error rate was increased by 9.2% per training session (Kruskal–Wallis and Dunnett's test, $P < 0.01$; Fig. 4b), whereas normal air-puff stimulation (1032 ± 593) had no significant effect. However, if the change in error rate was adjusted to account for the amount of feedback that the animals had received—calculated as $\sum[\Delta(\text{error rate}) \times N_{\text{twitches}}] / \sum N_{\text{twitches}}$ —there was a tendency for normal air-puff stimulation to reduce the error rate (Wilcoxon signed-rank test, $P < 0.087$). Uncorrelated air-puff stimulation (1159 ± 833) had no effect on NWR adaptation (Wilcoxon signed-rank test, $P = 0.65$; Fig. 4b). Thus, the behavioural data prove that NWR adaptation can indeed be guided by spontaneous muscle twitches.

The simulations show that the postulated reflex network is able to self-organize, guided solely by local learning rules without defined error signals. Each module extracts the relevant feedback information efficiently by assuming a 'learning mode' only in conjunction with its own twitches, given limited co-activity (see Methods and Supplementary Information). A system performance that asymptotically approaches zero error is achieved by multiple learn-

ing cycles, where each cycle causes a small change in synaptic weights. These learning principles result in a much lower error rate than classical hebbian learning in a feedforward network architecture, the latter often being assumed to underlie the spatial organization of sensory representations²³. The behavioural experiments also confirm that tactile input can be used to adjust the connection strengths functionally in the nociceptive spinal withdrawal reflex circuits⁸, but only when tactile input is correlated with spontaneous twitches. The stronger effect of aberrant as opposed to normal feedback in the behavioural study (Fig. 4b) was predicted by simulation of MDSI using aberrant and normal feedback (see Supplementary Information.).

In many sensory systems, spontaneous activity is known to be essential for the refinement of topographical representations: the projections of patches of synchronously discharging afferent neurons are strengthened centrally and those of unsynchronized (uncorrelated) inputs are weakened, thereby refining initial topographical representations that were relatively crude²⁴. Note, however, that the ensuing pattern of afferent input on spontaneous muscle twitches provides more complex functional information, integrating aspects of the complex body-surface, biomechanics and movement patterns. To our knowledge, MDSI has not been established in any other sensorimotor system, although it has been shown that active tactile exploration is needed for the maturation of normal vibrissa representation within the rat somatosensory system²⁵ (see also ref. 26).

The mechanisms that initiate the muscle twitches, and thus MDSI, in the reflex circuits are not known. However, cells located in the deep dorsal horn, some of which are presumably reflex encoders, exhibit Ca²⁺-dependent plateau potentials that greatly increase firing frequency and the duration of after-discharges²⁷. These plateau potentials might drive the spontaneous twitches and at the same time switch the membrane properties of the reflex interneurons from a 'transmission mode' to a 'learning mode'. Interestingly, spontaneous muscle twitches occur preferentially during REM-like sleep^{1,2,11}, indicating that MDSI is influenced by the brain's state. This would be consistent with the concept that the brain has a permissive role in this learning¹⁶. Sleeping and dreaming are known to promote learning²⁸ and also influence the functional development of sensory systems in young animals²⁹. The notion of using test pulses to probe the biomechanical properties of the system bears similarities to impulse response analysis, which is used abundantly to characterize linear time-invariant systems, for example in control theory³⁰. MDSI in self-learning systems could potentially be of great value in applications related to the design of prosthetics and in robotics.

The present results, which identify the role of spontaneous

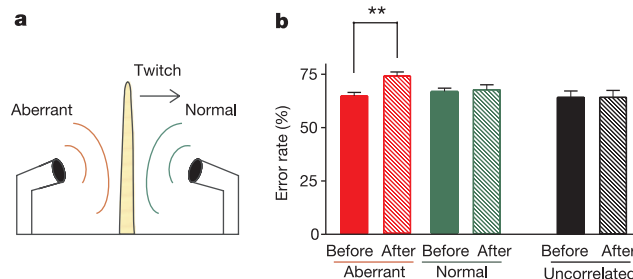


Figure 4 Behavioural experiments. **a**, Schematic training set-up depicting the aberrant and normal training protocols. **b**, NWR error rate before and after aberrant, normal or uncorrelated stimulation; results are means $\pm$ s.e.m. The NWR error rate was significantly increased (** $P < 0.01$) in rats given aberrant tactile feedback (49 training

sessions); normal air-puff stimulation (54 training sessions) had no significant effect (Kruskal–Wallis and Dunnett's test). Controls with random air-puff stimulation (23 training sessions) or no stimulation (23 training sessions; not shown) did not affect NWR adaptation ($P = 0.65$ and $P = 0.37$, respectively; Wilcoxon signed-rank test).

movements in the functional adaptation of spinal sensorimotor systems, provide a key to the understanding of how these basic circuits can incorporate information about the individual's bodily constitution and thereby perform the remarkably complex computations needed for accurate motor output. Many sensorimotor systems could have this strategy in common. □

Methods

Computer simulation and model parameters

Simulations were implemented in MATLAB (MathWorks; see Supplementary Information for the program code). Movement patterns produced by the activation of five different muscles (peroneus longus (PL), peroneus brevis (PB), tibialis anterior (TA), gastrocnemius (G) and two subunits of the extensor digitorum longus with tendon insertions on digits 2 and 3 (EDL23) and digits 4 and 5 (EDL45), respectively) were represented in 28×61 matrices (597 of these matrix elements represented sites of the plantar skin and were used in the simulation, representing >1.5 pixels mm^{-2} in an adult). The data were sampled by optical three-dimensional movement analysis yielding the velocity component perpendicular to the skin surface for 40–50 individual skin sites. Matrix values for intermediate sites were interpolated by spatial low-pass filtering⁶. These patterns were assumed to describe the pressure changes on the plantar skin in the hindlimb in standing position and consequently the pattern of afferent input from mechanoreceptors responding to skin pressure on muscle contraction. Quantified and normalized NWR receptive fields (intramuscular electromyographic recordings) of the muscles were obtained from previous studies⁶. For each learning epoch, muscles were activated at random with a specified probability. On co-activation of muscles, the withdrawal movement was approximated as the sum of the normalized withdrawal patterns of the individual activated muscles. The possible error in correlation coefficient caused by simple summation of movement patterns on co-activation was estimated to be less than 3.5% for a twitch probability per muscle of 0.045 (as calculated by using only the contaminating movement patterns on co-activation). The cutaneous feedback, which was represented on the same coordinates as the withdrawal efficiency, is conveyed by parallel afferents reaching 597 first-order interneurons in substantia gelatinosa (SG_j). These, in turn, act on six reflex-encoding deep dorsal-horn neurons (RE^k). Hence, 597 connection weights (w_j^k) between substantia gelatinosa neurons and reflex encoders for each of the six modules were updated in each learning epoch. A total of 10,000 epochs was estimated to correspond roughly to the observed number of tail twitches during the adaptation period¹⁴. Oja's rule, $\Delta w_j^k = \eta RE^k (SG_j - RE^k w_j^k)$, where η is the learning rate, was used for synaptic weight adjustments (w_j^k normalizes to $w_j^k \in [0, 1]$)^{19,20,23,24}. In MDSI, a learning epoch is initiated by a burst in the reflex-encoding neuron ($RE^k \neq 0$). Parameter values in the simulation shown in Fig. 3a, c had the following settings: a learning rate of 0.008, a twitch probability per module for each epoch of 0.045 (resulting in $\sim 1/4$ of the epochs containing twitch activity in at least one of the six modules and $\sim 11\%$ co-activity), sensory noise $\in [-0.5, 0.5]$ ($\pm 50\%$ of maximum tactile input), and initial synaptic weights were randomly drawn from a uniform distribution $U[-0.8, 0.8]$. The model was effective within a relatively wide range of parameter values. Mean adult correlations of at least 0.8 were reached for $\eta \in [0.0053, 0.043]$, twitching probability $\in [0.03, 0.24]$ (yielding 7.5–55% co-activity), sensory noise $\in [-3.9, 3.9]$ and initial synaptic weights drawn from $U[-3.2, 3.2]$ if these parameters were altered one at a time. Using only the contaminating movement patterns on co-activation, 0.11 is the maximum twitch probability that still satisfies a mean adult correlation of more than 0.80.

In the simulation of unsupervised feedforward learning, learning occurs when the postsynaptic activity in the reflex-encoding neurons exceeds $\sim 15\%$ of the maximum response. Increasing the activation probability did not improve adult correlations. When activated, weight adjustments were performed exactly as in MDSI learning.

Animals used

Wistar rats (P12–P17) of both sexes ($n = 84$ from 13 litters) were used.

Conditioning stimulation

Spontaneous twitches were detected by an optical system consisting of a complementary metal-oxide semiconductor sensor camera (CCi4; C-Cam Technologies, acquisition rate ~ 180 images s^{-1}) with software written in LabVIEW 5.2 (National Instruments Corp.) on a PC. Twitches with a lateral deviation of the midsection of the tail more than 9.2 radians s^{-1} (~ 0.28 m s^{-1}) after more than 300 ms of immobility triggered an air-puff directed to the side of the tail (Fig. 4a). These criteria were selected after a careful qualitative evaluation of the system's classification of different movement patterns. The air-puff had a peak pressure of ~ 20 kPa and stimulation latency from twitch initiation of ~ 40 ms at normal tail position. The time-point and direction of stimulation were stored by the system. Conditioning stimulation was given for 2–3 days per animal. The uncorrelated stimulation was given at random to either side of the tail. Trained animals did not differ from untrained animals of the same litter with regard to body weight or overall behaviour.

NWR testing

Every test included 10 stimulations (25-ms CO_2 -laser pulses, 5 W, beam width 3 mm) to each side of the distal tail; the persons performing the analyses were blinded to the type of conditioning stimulation used. Response latencies were monitored in all NWR testing experiments; owing to the increase in animal size, the mean latencies increased from 156 to 179 ms during the P12–P17 period (mean of the s.e.m. = 2.1 ms), indicating C-fibre-

evoked reflexes. Only tail-tip movements of more than 0.02 radians (~ 1 mm) were considered. Data were discarded from the study (53 sessions) if the error rate was less than 20% or more than 80% before conditioning, to allow the detection of both improvements and deteriorations (mean of the s.d. within a given testing session $\approx 13\%$).

Received 17 January; accepted 23 April 2003; doi:10.1038/nature01719.

- Karlsson, K. A. & Blumberg, M. S. The union of the state: Myoclonic twitching is coupled with nuchal muscle atonia in infant rats. *Behav. Neurosci.* **116**, 912–917 (2002).
- Blumberg, M. S. & Lucas, D. E. A developmental and component analysis of active sleep. *Dev. Psychobiol.* **29**, 1–22 (1996).
- Blumberg, M. S. & Lucas, D. E. Dual mechanisms of twitching during sleep in neonatal rats. *Behav. Neurosci.* **108**, 1196–1202 (1994).
- Clancy, B., Darlington, R. B. & Finlay, B. L. Translating developmental time across mammalian species. *Neuroscience* **105**, 7–17 (2001).
- Pouget, A. & Snyder, L. H. Computational approaches to sensorimotor transformations. *Nature Neurosci.* **3** (Suppl.), 1192–1198 (2000).
- Schouenborg, J. & Weng, H. R. Sensorimotor transformation in a spinal motor system. *Exp. Brain Res.* **100**, 170–174 (1994).
- Holmberg, H. & Schouenborg, J. Postnatal development of the nociceptive withdrawal reflexes in the rat: a behavioural and electromyographic study. *J. Physiol. (Lond.)* **493**, 239–252 (1996).
- Waldenström, A., Thelin, J. & Schouenborg, J. Tactile sensory input is used for the postnatal tuning of the nociceptive withdrawal reflex system. *Soc. Neurosci. Abstr.* **30**, 1623 (2001).
- Holmberg, H. & Schouenborg, J. Developmental adaptation of withdrawal reflexes to early alteration of peripheral innervation in the rat. *J. Physiol. (Lond.)* **495**, 399–409 (1996).
- Holmberg, H., Schouenborg, J., Yu, Y. B. & Weng, H. R. Developmental adaptation of rat nociceptive withdrawal reflexes after neonatal tendon transfer. *J. Neurosci.* **17**, 2071–2078 (1997).
- Gardner, R. & Grossman, W. Normal motor patterns in sleep in man. *Adv. Sleep Res.* **2**, 67–107 (1975).
- Hadders-Algra, M., Nakae, Y., Van Eykern, L. A., Klip-Van den Nieuwendijk, A. W. & Precht, H. F. The effect of behavioural state on general movements in healthy full-term newborns. A polymyographic study. *Early Hum. Dev.* **35**, 63–79 (1993).
- de Lisi, L. Su di un fenomeno motorio costante del sonno normale: Le mioclonie ipniche fisiologiche. *Riv. Patol. Nerv. Ment.* **29**, 481–496 (1932).
- Waldenström, A., Christensson, M. & Schouenborg, J. Spontaneous movements precede and overlap in time with the tuning of the nociceptive withdrawal reflex (NWR) in postnatal rats. *IASP Abstr.* **1558**, P106 (2002).
- Fitzgerald, M. & Koltzenburg, M. The functional development of descending inhibitory pathways in the dorsolateral funiculus of the newborn rat spinal cord. *Brain Res.* **389**, 261–270 (1986).
- Levinsson, A., Luo, X. L., Holmberg, H. & Schouenborg, J. Developmental tuning in a spinal nociceptive system: effects of neonatal spinalization. *J. Neurosci.* **19**, 10397–10403 (1999).
- Cervero, F. & Iggo, A. The substantia gelatinosa of the spinal cord: a critical review. *Brain* **103**, 717–772 (1980).
- Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386–408 (1958).
- Oja, E. A simplified neuron model as a principal component analyzer. *J. Math. Biol.* **15**, 267–273 (1982).
- Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C. & Nelson, S. B. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature* **391**, 892–896 (1998).
- Nakatsuka, T., Ataka, T., Kumamoto, E., Tamaki, T. & Yoshimura, M. Alteration in synaptic inputs through C-afferent fibers to substantia gelatinosa neurons of the rat spinal dorsal horn during postnatal development. *Neuroscience* **99**, 549–556 (2000).
- Linsker, R. From basic network principles to neural architecture: emergence of orientation columns. *Proc. Natl Acad. Sci. USA* **83**, 8779–8783 (1986).
- Fregnac, Y. & Bienenstock, E. in *Mechanistic Relationships between Development and Learning* (eds Carew, T. J., Menzel, R. & Shatz, C. J.) 113–148 (Wiley, Berlin, 1998).
- Katz, L. C. & Shatz, C. J. Synaptic activity and the construction of cortical circuits. *Science* **274**, 1133–1138 (1996).
- Nicoletis, M. A., De Oliveira, L. M., Lin, R. C. & Chapin, J. K. Active tactile exploration influences the functional maturation of the somatosensory system. *J. Neurophysiol.* **75**, 2192–2196 (1996).
- Levinsson, A., Holmberg, H., Broman, J., Zhang, M. & Schouenborg, J. Spinal sensorimotor transformation: Relation between cutaneous somatotopy and a reflex network. *J. Neurosci.* **22**, 8170–8182 (2002).
- Morisset, V. & Nagy, F. Nociceptive integration in the rat spinal cord: Role of non-linear membrane properties of deep dorsal horn neurons. *Eur. J. Neurosci.* **10**, 3642–3652 (1998).
- Stickgold, R., Hobson, J. A., Fosse, R. & Fosse, M. Sleep, learning, and dreams: Off-line memory reprocessing. *Science* **294**, 1052–1057 (2001).
- Frank, M. G., Issa, N. P. & Stryker, M. P. Sleep enhances plasticity in the developing visual cortex. *Neuron* **30**, 275–287 (2001).
- Johansson, R. *System Modeling and Identification* (Prentice Hall, Englewood Cliffs, New Jersey, 1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Nockhammar for technical assistance and M. Garwicz for constructive comments on earlier versions of the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.P. (per.petersson@mphy.lu.se).

Melanopsin and rod–cone photoreceptive systems account for all major accessory visual functions in mice

S. Hattar^{*}, R. J. Lucas[†], N. Mrosovsky[‡], S. Thompson[†], R. H. Douglas[§], M. W. Hankins[†], J. Lem^{||}, M. Biele[¶], F. Hofmann[#], R. G. Foster[†] & K.-W. Yau^{*}

^{*}Howard Hughes Medical Institute and Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

[†]Department of Integrative and Molecular Neuroscience, Division of Neuroscience and Psychological Medicine, Faculty of Medicine, Imperial College London, Charing Cross Campus, London W6 8RF, UK

[‡]Departments of Zoology, Physiology and Psychology, University of Toronto, Toronto, Ontario M5S 3G5, Canada

[§]Applied Vision Research Centre, Department of Optometry and Visual Science, City University, Northampton Square, London EC1V 0HB, UK

^{||}Department of Ophthalmology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA

[¶]Lehrstuhl Pharmakologie für Naturwissenschaften, Zentrum für Pharmaforschung, Ludwig-Maximilians Universität München, 81377 München, Germany

[#]Institut für Pharmakologie und Toxikologie, Technische Universität München, 80802 München, Germany

In the mammalian retina, besides the conventional rod–cone system, a melanopsin-associated photoreceptive system exists that conveys photic information for accessory visual functions such as pupillary light reflex and circadian photo-entrainment^{1–7}. On ablation of the melanopsin gene, retinal ganglion cells that normally express melanopsin are no longer intrinsically photosensitive⁸. Furthermore, pupil reflex⁸, light-induced phase delays of the circadian clock^{9,10} and period lengthening of the circadian rhythm in constant light^{9,10} are all partially impaired. Here, we investigated whether additional photoreceptive systems participate in these responses. Using mice lacking rods and cones, we measured the action spectrum for phase-shifting the circadian rhythm of locomotor behaviour. This spectrum matches that for the pupillary light reflex in mice of the same genotype¹¹, and that for the intrinsic photosensitivity of the melanopsin-expressing retinal ganglion cells⁷. We have also generated mice lacking melanopsin coupled with disabled rod and cone photo-transduction mechanisms. These animals have an intact retina but fail to show any significant pupil reflex, to entrain to light/dark cycles, and to show any masking response to light. Thus, the

rod–cone and melanopsin systems together seem to provide all of the photic input for these accessory visual functions.

Rods and cones have long been thought to be the exclusive photoreceptors in the retina. This hypothesis is now known to be untrue. An opsin-like protein called melanopsin, originally identified in *Xenopus* skin melanophores¹, is present in a small subset of mammalian retinal ganglion cells (RGCs)^{1–6}, and these cells are intrinsically photosensitive^{6,7}. The axons of these RGCs project predominantly to the suprachiasmatic nucleus (SCN), the intergeniculate leaflet (IGL) and the olivary pretectal nucleus (OPN) of the brain⁶, which are key centres for circadian photo-entrainment and the pupillary light reflex. In melanopsin knockout mice (*Opn4*^{−/−}, formerly referred to as *mop*^{−/−}; see ref. 8), those RGCs that would normally express melanopsin lose their intrinsic photosensitivity⁸. *Opn4*^{−/−} mice also have an incomplete pupillary light reflex at high illuminations⁸. In independently produced melanopsin-knockout mice, others have found that the ability of light to phase-delay and lengthen the period of the circadian rhythm is also diminished^{9,10}. For the pupil reflex, this photic response can be quantitatively accounted for by a functional complementarity between the rod–cone system and the melanopsin system, without the need to invoke any additional light-detection system⁸. Nonetheless, the proposal has persisted that cryptochromes—flavoproteins reported to have a direct light-detecting role in *Drosophila*^{12,13}—may have the same function in mammals^{14–16} despite earlier evidence to the contrary¹⁷. To settle this question, we first examined the action spectrum for phase-shifting the circadian rhythm in mice lacking rod and cone photoreceptors (*rd/rd cl*)¹⁸. Next, we generated triple-knockout mice lacking all confirmed photodetection systems—*Opn4*^{−/−} *Gnat1*^{−/−} *Cnga3*^{−/−} (melanopsin (also known as opsin 4), guanine nucleotide-binding protein α -transducin 1 (also known as rod transducin α -subunit, or Tr α) and cyclic GMP-gated channel A-subunit 3, respectively)—and tested these animals for pupil reflex, circadian photo-entrainment and the masking response to light.

Irradiance–response relations for the light-induced phase shifting of the circadian rhythm of locomotor behaviour in *rd/rd cl* mice were measured at various wavelengths (Fig. 1a). The irradiance for half-maximal phase shift at each wavelength was then plotted to give the action spectrum (Fig. 1b). This spectrum is best fitted by the predicted absorption spectrum of a vitamin A₁-based photopigment with a wavelength of maximum absorbance ($\lambda_{\max}$) $\approx$ 481 nm, similar to that for the pupil reflex in this genotype ($\lambda_{\max} \approx$ 479 nm)¹¹ and, more specifically, that for the intrinsically photosensitive, melanopsin-expressing RGCs in the rat ($\lambda_{\max} \approx$ 484 nm)⁷. Thus in the absence of rods and cones, circadian photo-entrainment is apparently determined by the melanopsin system. Previously, an action spectrum with a $\lambda_{\max}$ of 480 nm has been reported for

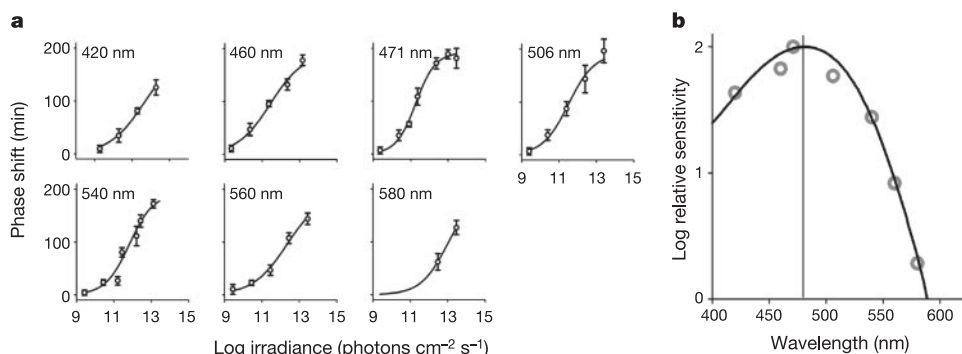


Figure 1 Irradiance–response relations (**a**) and action spectrum (**b**) for circadian phase shifting in *rd/rd cl* mice by monochromatic light between 420 nm and 580 nm, assayed by wheel-running ($n = 4–7$ animals per irradiance at each wavelength). The derived action

spectrum in **b** for circadian photo-entrainment is best fitted ($R^2 = 0.976$) by the absorption spectrum of a vitamin A₁-based photopigment with $\lambda_{\max} = 481$ nm (continuous curve).

circadian phase shifting by light in one strain of *rd* mice that has lost rods and perhaps most cones³¹. However, in part because the same experiments on a different strain of *rd* mice gave a $\lambda_{\max}$ between 500 nm and 515 nm (ref. 32) and interpreted to reflect the action spectrum of mouse middle-wavelength-sensitive (M)-cones, the significance of the first study was never settled. With the *rd/rd cl* mouse line described here in which the loss of rods and cones is essentially complete, however, the signal from residual cones is no longer an issue.

To obtain more conclusive evidence that no other independent photoreceptive systems exist for the various accessory visual functions, we generated triple-knockout mice in which the rod-cone system and the melanopsin system were both silenced. Notably, the silencing of rod-cone functions in this case was achieved not by inducing degeneration of these cells (as occurs in *rd/rd cl* mice) but by combining targeted deletions of the genes for rod *Gnat1* (ref. 19) and cone *Cnga3* (refs 20, 21). *Gnat1* and *Cnga3* are critically involved in the G-protein-coupled cGMP signalling pathway that

mediates rod-cone phototransductions. Therefore, in triple-knockout mice, the melanopsin system and the rod-cone system are both unable to signal light. We investigated whether these animals still had any residual response to light, as assayed by their pupil reflex and locomotor behaviour.

The *Opn4*^{-/-} *Gnat1*^{-/-} *Cnga3*^{-/-} mice were generated by first producing triple heterozygotes (*Opn4*^{+/-} *Gnat1*^{+/-} *Cnga3*^{+/-}) and then mating these to homozygosity. These mice had a normal-looking retina (Fig. 2a). Because *Opn4*^{-/-} was produced by replacing the melanopsin gene with the tau-LacZ construct^{6,8}, the RGCs that would normally express melanopsin could be visualized by 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal) labelling.

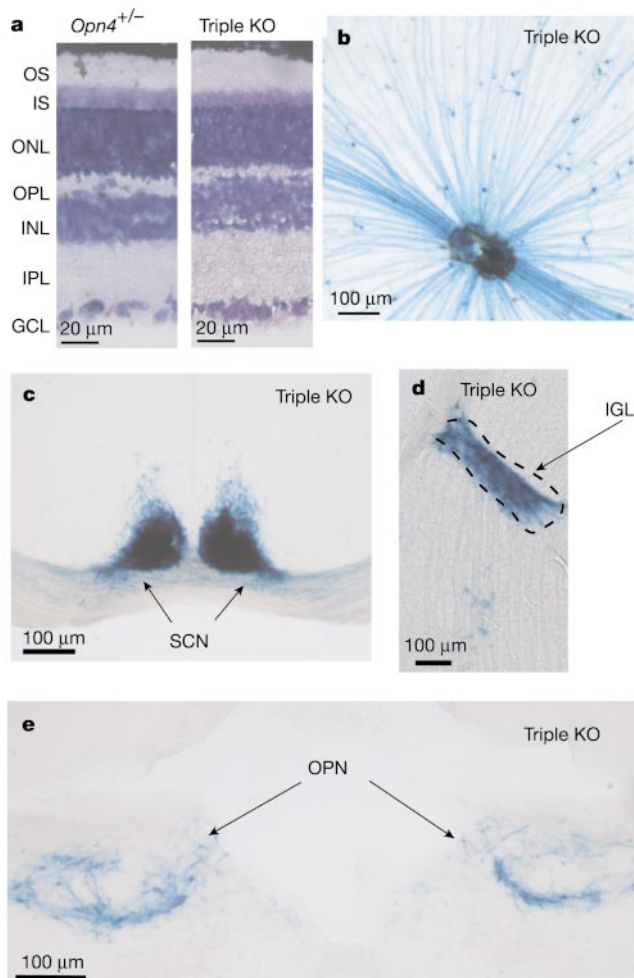


Figure 2 Normal retinal morphology and presence and central connectivity of melanopsin-expressing RGCs in triple-knockout (*Opn4*^{-/-} *Gnat1*^{-/-} *Cnga3*^{-/-}) mice. **a**, Retinal cross-sections from *Opn4*^{+/-} and triple-knockout (KO) mice. Giemsa stain shows the various layers. Both are similar in morphology and thickness to wild type. GCL, ganglion cell layer; INL, inner nuclear layer; IPL, inner plexiform layer; IS, inner segment layer; ONL, outer nuclear layer; OPL, outer plexiform layer; OS, outer segment layer. **b**, Flat-mount view of a triple-knockout retina stained with X-gal (blue). **c–e**, Coronal sections of triple-knockout mouse brain showing the normal innervations of the SCN, IGL and OPN by X-gal-labelled axons. Dorsal side is up in each case.

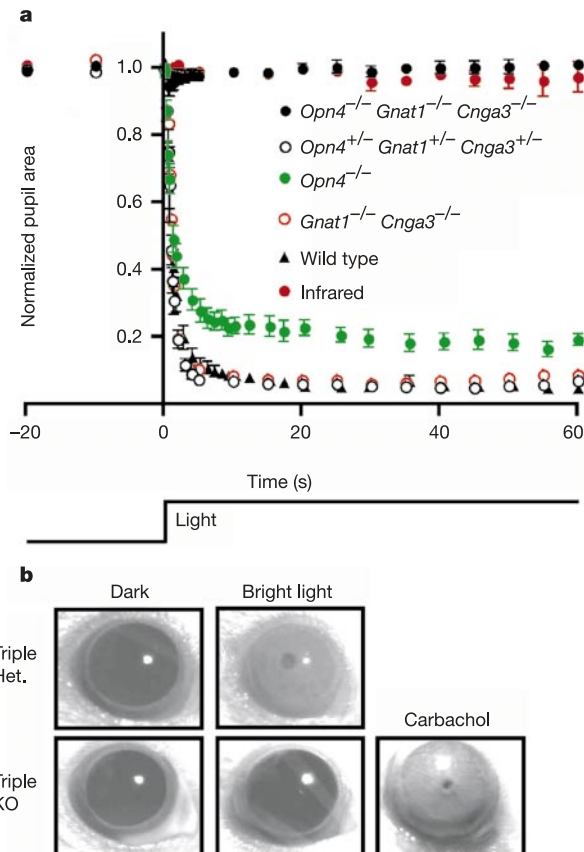


Figure 3 Disabling of rods, cones and melanopsin-positive RGCs essentially eliminates the pupillary light reflex. **a**, Pupil area (mean $\pm$ s.e.m.) as a percentage of dark-adapted aperture area (recorded just before light exposure, time 0 on graph) for *Opn4*^{-/-} *Gnat1*^{-/-} *Cnga3*^{-/-} ($n = 6$), *Opn4*^{+/-} *Gnat1*^{+/-} *Cnga3*^{+/-} ($n = 4$), *Opn4*^{-/-} ($n = 8$; data reproduced from ref. 8) and *Gnat1*^{-/-} *Cnga3*^{-/-} ($n = 1$) mice over the course of 1 min when exposed to 480-nm light at 86–140 μ W cm⁻². The wild-type data are from ref. 8. For comparison, the pupil size of triple-knockout mice exposed to an infrared light pulse ($n = 6$) is also shown. On close scrutiny, a small, transient pupil constriction was detected in two out of six triple-knockout mice exposed to the intense 480-nm stimulus. The averaged 480-nm data shown for the triple knockouts are from mice dark-adapted for 3 days before light exposure, but are representative of similar recordings from the same group after 1-h dark adaptation. Neither the amplitude nor the frequency of occurrence of the residual pupillary response in the triple-knockout animals were increased with bright white-light stimulus (30 mW cm⁻²), or with monochromatic stimuli ranging from 360 to 660 nm (70 μ W cm⁻² at 360 nm; 19 μ W cm⁻² at 420 nm; 86 μ W cm⁻² at 480 nm; 100 μ W cm⁻² at 550 nm; 68 μ W cm⁻² at 660 nm). Nor did these parameters show any significant change when tested repeatedly over a 24-h period. **b**, Application of 100-mM carbachol resulted in a complete pupil constriction in triple-knockout mice.

In the triple-knockout mice, these cells were still present (Fig. 2b) in numbers (about 600 per retina) comparable to wild type^{6,8}, and their axons still projected predominantly to the SCN, IGL and OPN of the brain (Fig. 2c–e). Thus, the absence of melanopsin and of functional rod–cone phototransductions does not affect the genesis, survival and central connectivity of the melanopsin-associated RGCs.

When tested for the pupillary light reflex with steady, intense exposure of light at 480 nm, the triple-heterozygous mice gave a response resembling wild type in amplitude and time course (Fig. 3a). The double-knockout mice (*Gnat1*^{−/−} *Cnga3*^{−/−}) showed the same response (Fig. 3a), similar to what we found previously¹¹ for *rd/rd cl* mice, which have complete degeneration of rods and cones. Triple knockouts, however, hardly gave any response to the 480-nm stimulus (Fig. 3a), nor to monochromatic light at other wavelengths (360–660 nm) or intense white light (data not shown). The same general results were obtained when these animals were tested at various times over a 24-h period. On close inspection, there was a very transient, barely detectable pupil response (with a mean of 5% reduction in pupil area at peak response) in two out of six tested triple-knockout mice under exposure to bright 480-nm light. However, even in animals where this residual response was detected, it was not consistently present on repeated stimulus trials with extensive dark adaptation (up to 3 days) in between (see Fig. 3 legend). The response was unlikely to be caused by heat associated with the illumination, because it largely disappeared after replacement of the 480-nm interference filter with a band-pass filter transmitting only infrared light (≥ 850 nm) (Fig. 3a). A possible source of the signal is the early receptor potential of rods and cones, which consists of a very small, transient membrane hyperpolarization caused by charge movements associated with conformational changes of the visual pigments after photo-isomerization^{22,23}. This hyperpolarization should persist regardless of whether the transduction steps downstream of the pigment are disabled or not. Another possibility is a very small, transient hyperpolarization generated by intense light in rods and cones via a pathway apparently independent of transducin²⁴. In any

case, the smallness of the residual pupil response, its transient nature even in steady, intense light, and the inconsistency of its occurrence all suggest that it is unlikely to be of physiological significance. Finally, the failure of the *Opn4*^{−/−} *Gnat1*^{−/−} *Cnga3*^{−/−} pupil to constrict was not due to a defect in the constriction mechanism because carbachol, a parasympathetic agonist, was able to activate maximum pupil constriction when applied to the cornea (Fig. 3b).

To assess photo-entrainment, *Opn4*^{+/−} *Gnat1*^{+/−} *Cnga3*^{+/−} and *Opn4*^{−/−} *Gnat1*^{−/−} *Cnga3*^{−/−} animals separate from those used in the pupil-reflex studies were kept in an 16/8-h light/dark cycle (800 lx white light in the light period), and their locomotor activity was monitored (see Methods). The triple-heterozygous mice showed normal photo-entrainment, whereas the triple-knockout mice did not show any entrainment (Fig. 4; see also Supplementary Information). Actograms indicated that the triple-heterozygous mice had an average period length of 24.0 ± 0 h (mean $\pm$ s.e.m., $n = 6$; individually all 24.0 h), as expected from stable photo-entrainment. In contrast, the triple homozygous mice had a period length of 23.3 ± 0.2 h ($n = 5$; individually 23.8, 23.3, 23.2, 22.5 and 23.5 h), similar to the 23.6 h reported for the same strain (B6/129) of wild-type mice in constant darkness²⁵, indicating that these animals free-ran even under light/dark conditions. Of note, at the light intensity used in these experiments, the pupils of wild-type or triple-heterozygous mice would have constricted considerably (see Fig. 3a), whereas those of the triple-knockout mice would have stayed unchanged. Thus, the equivalent light intensity at which the triple-knockout mice failed to show photo-entrainment was actually much higher than that capable of entraining wild-type or triple-heterozygous mice.

We have also examined another accessory visual function, namely the masking of locomotion of nocturnal rodents by light, which involves a fast and direct effect of light independent of the circadian pacemaker^{26–28}. To examine masking by light, the mice tested for photo-entrainment were subsequently placed on a 3.5/3.5-h light/dark cycle (800 lx white light as before, although the equivalent intensity for the triple-knockout mice would again be considerably

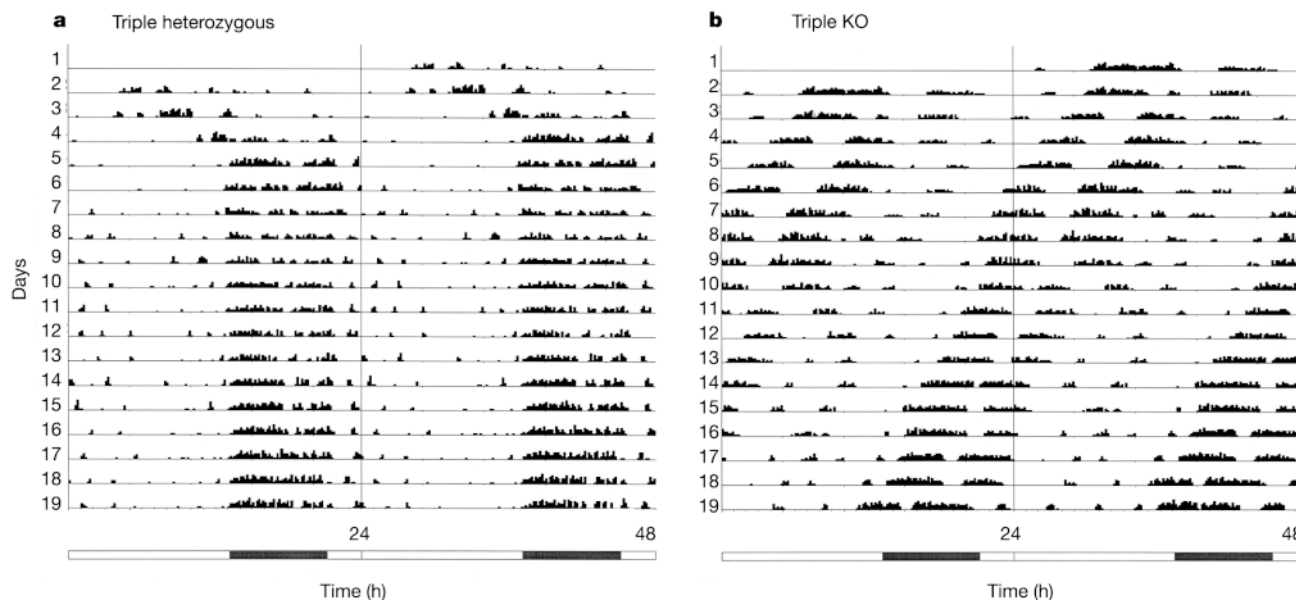


Figure 4 Actograms of wheel running for mice under a 16/8-h light/dark cycle, double-plotted on a 24-h timescale. The illumination was approximately 800 lx white light. The numbered lines represent successive days. Activity levels (in total number of revolutions in 10-min bins) are given in 15 quantiles, with the first being 1–55 revolutions, the second 56–110, and so on. The bar below the actograms indicates light (white) and dark (black)

periods. **a**, Triple heterozygote (*Opn4*^{+/−} *Gnat1*^{+/−} *Cnga3*^{+/−}). The locomotor activity had a cycle very close to 24 h and was phase-locked to darkness, showing photo-entrainment. **b**, Triple-knockout (*Opn4*^{−/−} *Gnat1*^{−/−} *Cnga3*^{−/−}). The animal free-ran with a period of less than 24 h, showing lack of photo-entrainment.

higher than for the triple-heterozygous mice when the lack of pupil reflex in the triple-knockout animals is taken into account). This ultradian cycle is useful for assessing masking because it is difficult to entrain circadian rhythms to cycles with periods of 7 h or multiples thereof. In this way, it was ensured that masking could be measured at different phases of the endogenous clock²⁹. When kept in this cycle, the triple-heterozygous mice showed two effects. Four of the six tested animals were negatively masked (that is, their locomotor activity was diminished) by light (Fig. 5a). The average percentage of activity in the dark period for these 'light avoiders' was $87.1 \pm 8.8\%$ (mean $\pm$ s.e.m.; individually 99.5%, 66.3%, 85.0% and 97.7%), similar to what was seen with B6/129 wild-type mice ($n = 12$; N.M. and S.H., unpublished data). The remaining two animals, paradoxically, developed more wheel-running activity in the light period (Fig. 5b), possibly due to the presence of the 129-strain background associated with the triple-heterozygous genotype (this activity reversal has previously been observed in other mouse lines with the 129-strain background; R.J.L., unpublished data). Although the reason for this reversal of activity with respect to the light/dark cycle is unclear, undoubtedly light still had an effect (in this case, positive masking). The percentage of activity in the dark period for these two 'light preferers' was 32.7% (Fig. 5b) and 26.0%, respectively. The triple-knockout mice, on the other hand, were hardly affected by this ultradian light/dark cycle (Fig. 5c, d). The average percentage of activity in the dark period was $43.2 \pm 3.4\%$ (mean $\pm$ s.e.m., $n = 4$; individually 44.7%, 50.5%, 40.8% and 36.6%), much closer to that expected from randomness (50%).

Given the fact that RGCs normally expressing melanopsin still project to the appropriate brain targets in triple-knockout mice, the conclusion from our experiments is that the rod-cone system and the melanopsin system are the exclusive light-detecting systems in the eye for producing the normal pupillary light reflex, photo-entrainment and masking response to light. These experiments go beyond our previous conclusion that the rod-cone and melanopsin systems are fully complementary to each other in the pupil reflex⁸, by demonstrating an essentially complete loss of this and other functions when the two systems are absent or disabled. It was recently reported¹⁶ that the pupil reflex of mice lacking rods as well as Cry1 and Cry2 (*rd/rd Cry1^{-/-} Cry2^{-/-}*) is 100-fold less photosensitive than normal (and 10-fold less sensitive than mice only lacking rods; *rd/rd*), proposed to support the idea that the cryptochromes may participate as a photodetective system. Nonetheless, other interpretations of these earlier data are possible. Furthermore, the pupil reflex of mice only lacking Cry1 and Cry2 is normal (as is their masking response to light³⁰). A potential complicating factor in interpreting the results from mice lacking Cry1 and Cry2 is that these animals are reported to show frequent ocular inflammation¹⁵, which may cause subtle changes in retinal function. In any case, we know from immunocytochemistry of our *Opn4^{-/-} Gnat1^{-/-} Cnga3^{-/-}* mice that at least Cry2 is still widely expressed in the inner retina (including the RGCs expressing melanopsin), comparable to wild type (data not shown)—the antibodies for Cry1 did not give informative labelling; see Methods. At the same time, real-time polymerase chain reaction with reverse transcription (RT-PCR) indicated that the message levels for Cry1

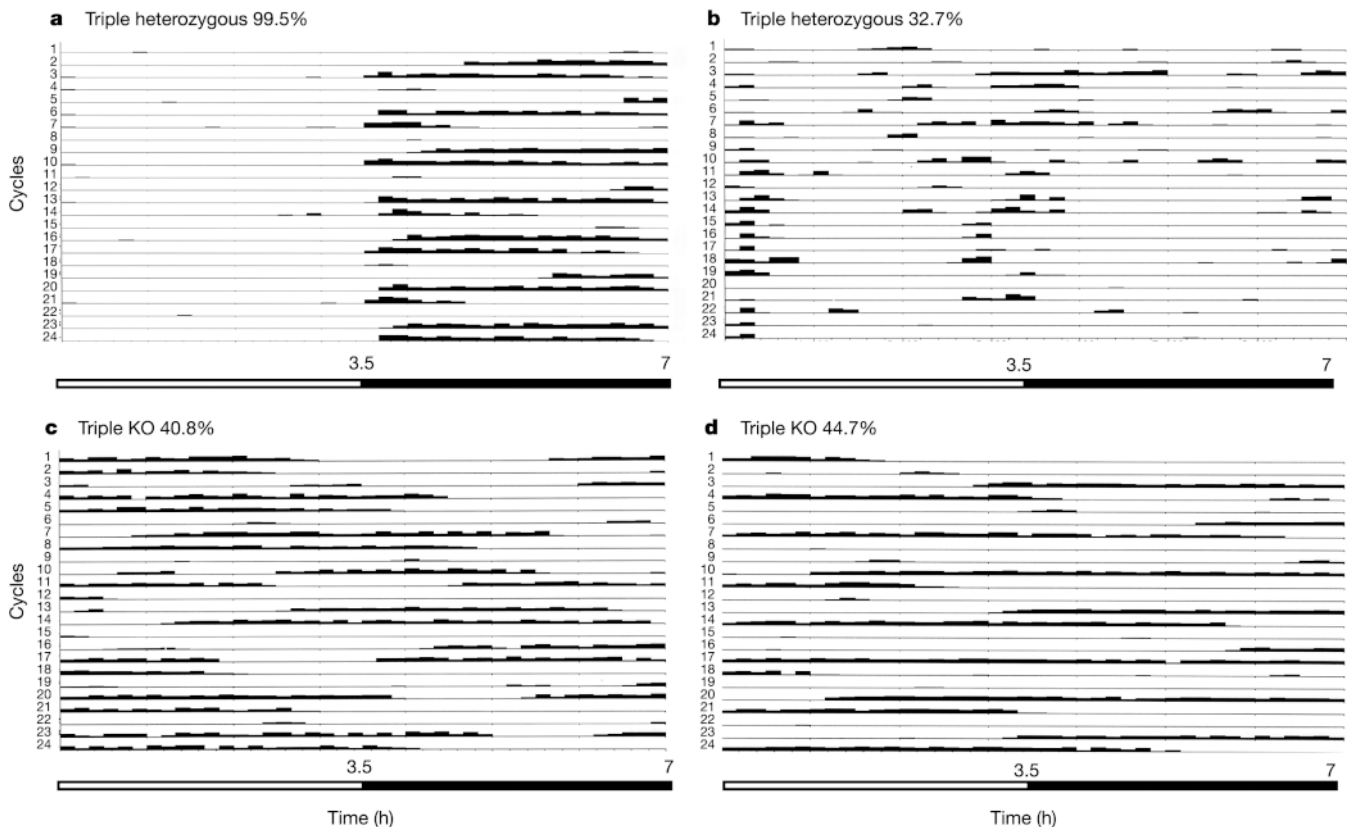


Figure 5 Actograms of wheel running plotted on a 7-h timescale. The mice were subjected to an ultradian 3.5/3.5-h light/dark cycle (see text). Illumination was approximately 800 lx white light. The numbered lines represent successive 7-h cycles. The activity quantiles and time bins are the same as in Fig. 4. **a, b**, Two triple-heterozygous mice (*Opn4^{+/-} Gnat1^{+/-} Cnga3^{+/-}*) showing a negative and positive

masking effect, respectively. The percentage of total revolutions in darkness was 99.5% and 32.7%, respectively. **b**, Two triple-knockout mice (*Opn4^{-/-} Gnat1^{-/-} Cnga3^{-/-}*). The percentage of total revolutions in darkness was 40.8% and 44.7%, respectively, not far from the 50% expected from randomness; that is, no masking effect of light.

and Cry2 in *Opn4*^{-/-} *Gnat1*^{-/-} *Cnga3*^{-/-} eyes are comparable to those in wild-type mice (Supplementary Information). Thus, the disappearance of the accessory visual functions in the triple-knockout mice cannot be attributed to secondary loss of cryptochromes. Even if mammalian cryptochromes should turn out to function as direct light detectors, either they obligatorily depend on the presence of melanopsin for their light-detecting function or they do not signal light at least for the diverse photic responses studied here. □

Methods

Animals

The *rd/rd cl* mice were bred at Imperial College London, and the circadian phase-shifting experiments were carried out there. The triple-knockout and the corresponding triple-heterozygous mice were bred at Johns Hopkins. The pupil reflex experiments were done at Hopkins, whereas the wheel-running experiments were done at the University of Toronto. At the time of experimentation, the Hopkins animals were 1–4 months old for the pupil reflex measurements and 1–9 months old (1–9 months for triple-heterozygous mice, 1–2 months for triple-knockout mice; all were males except for one female triple knockout, which was used for photo-entrainment but not for masking experiments) for the wheel-running measurements. For morphological, X-gal labelling, immunocytochemical and RT–PCR experiments, the animals were 2–3 months old.

Giemsa staining of retinal cross-sections

An animal anaesthetized by intraperitoneal injection of avertin (0.2 ml g⁻¹) was circulationally perfused with 25 ml cold 4% paraformaldehyde in PBS for 15 min. The eyes were isolated, corneas removed, and fixed for an additional 3 h in cold 4% paraformaldehyde in PBS. The eyes were cryo-protected overnight in 30% sucrose in PBS at 4°C, then sectioned at 10 µm thickness on a cryostat. The sections were allowed to dry and stained with the nucleus-labelling Giemsa solution (Sigma).

Assays for Cry1 and Cry2 in triple-knockout eyes

We used two assays. The first was immunocytochemistry⁶, with antibodies PA1-527 (Affinity Bioreagents) and CRY11-A (Alpha Diagnostic International) against Cry1, and CRY21-A (Alpha Diagnostic International) against Cry2. The two antibodies against Cry1 did not give specific labelling in the retina, but Cry2 gave labelling specific enough for comparison with wild-type retina. The second assay was with RT–PCR using Cry1- and Cry2-specific primers. The samples were whole eyes without the lens, and RNA was extracted using the trizol method (Sigma). The controls were RNA samples without reverse transcriptase. Real-time RT–PCR was carried out with SYBR green detection on an ABI Prism 7700 machine (Applied Biosystems). The identity of the product was checked by melting-curve analysis, size confirmation on an agarose gel and sequencing of the products.

X-gal staining

For X-gal labelling of the brain sections or flat-mount retina of the triple-knockout mice, an animal was anaesthetized and circulationally perfused with fixative as above, and the brain/eyes were isolated. An eye-cup was prepared by removing the anterior half of the eye, while the brain was frozen on dry ice and sectioned at 50-µm thickness with a sliding microtome (Leica SM2000R). Brain sections were mounted on microscope slides (Fisherbrand, superfrost-plus) and allowed to dry overnight. To stain for β-galactosidase activity, both preparations were washed twice in buffer B (100 mM phosphate buffer, pH 7.4, 2 mM MgCl₂, 0.01% Na-desoxycholate and 0.02% (octylphenoxy) polyethoxyethanol (IGEPAL)), and then incubated in the staining solution (buffer B plus 5 mM K-ferricyanide, 5 mM K-ferrocyanide and 1 mg ml⁻¹ X-gal) for 6–10 h at room temperature in darkness. Afterwards, the retina was isolated from the eye-cup and flat mounted. Both brain sections and the retina were mounted in glycerol. Images were obtained using a Zeiss Axiophot microscope and AxioCam HRC digital camera.

Phase shifting of locomotion in *rd/rd cl* mice

Singly housed male C3H/He *rd/rd cl* mice aged between 80 and 250 days were maintained in constant darkness. A single 15-min monochromatic light pulse (half bandwidth = 10 nm) of defined irradiance was applied after 7 days at circadian time 16 (4 h after activity onset) at 420, 460, 471, 506, 540, 560 and 580 nm, respectively. To correct for lens transmission at different wavelengths, lenses were extracted from *rd/rd cl* mice (*n* = 4) and wavelength-scanned in air using an integrating sphere (see Supplementary Information). The irradiances indicated have been corrected for lens transmission. The magnitude of phase shift in free-running locomotor activity rhythm by light at each wavelength was plotted against irradiance and fitted to a sigmoidal function. The irradiances that produced half-maximum phase shifts were plotted against wavelength to produce the action spectrum. This spectrum was fitted with the absorption spectrum of a vitamin A₁-based pigment with a best-fit λ_{max} determined by least squares.

Pupillometry

Consensual pupillary constriction was measured in response to an adirectional light stimulus. Light was provided by a PTI xenon arc-lamp, filtered to remove wavelengths <299 nm and most infrared, and transmitted by means of a liquid light guide to a diffusing sphere. This sphere was constructed from a ping-pong ball and designed to render the light stimulus nearly adirectional at the location of the mouse cornea. An adult

mouse, dark-adapted for 1 h and unanaesthetized, was positioned and oriented with the help of an infrared image-converter such that one eye was situated at the exit of the reflective sphere and the other eye was in focus, and oriented as perpendicular as possible, to a charge-coupled device (CCD) camera fitted with a × 10 macro lens. Illumination of the videoed eye was with infrared LEDs (>850 nm) throughout the experiment. Pupil measurements from the video images of the eye were made using TINA 3.1 image-analysis software and a Matrox meteor-II frame-grabber card (Matrox Imaging). The intensity and wavelength of the light stimulus were controlled with calibrated neutral-density and interference (10-nm bandwidth) filters. Irradiance measurements (W cm⁻²) were made with a radiometer (Macam Photometrics). Unless indicated otherwise, all experiments reported here were conducted during the light period of a 14/10-h light/dark cycle (between 9 and 3 h before lights off)⁸.

Carbachol application

Carbachol (carbamylcholine chloride) was obtained from Sigma. Solution (100 mM) was prepared in × 1 PBS, pH 7.4, and filtered through a 0.22-µm filter. Samples were topically applied to the cornea of the mouse eye with a microtip transfer pipette. The pupil size in darkness was assessed immediately before and at 5 min after carbachol application. We found that a carbachol concentration of 100 mM was able to fully constrict the eyes of both triple-heterozygous and triple-knockout mice, and therefore we did not use higher concentrations (see ref. 8).

Photo-entrainment and masking experiments

The mice were placed individually in cages (44 × 23 × 20 cm) equipped with running wheels (17.5 cm in diameter). Room temperature was 19.5–23.5°C. Revolutions were recorded with Dataquest III hardware and software in 10-min bins. The light/dark cycle was 16/8-h light/dark or 3.5/3.5-h light/dark, with approximately 800 lx in the light period (Hagner EX2 luxmeter) illuminated by Sylvania Octron 32 W 4100K fluorescent tubes. For photo-entrainment, the animals were studied for 19 days on the 16/8-h light/dark cycle, with data from the last 10 days used for period analysis based on the algorithm of Clocklab (Actimetrics) for the Bushell and Sokolove χ² periodogram. For masking, the animals were kept on the ultradian 3.5/3.5-h light/dark cycle for 11 days. Data from the last 7 days, which covered light and dark periods spanning the entire circadian cycle, were used for calculating the percentages of wheel-running activity in darkness.

Received 21 April; accepted 2 June 2003; doi:10.1038/nature01761.

Published online 15 June 2003.

- Provincio, I., Jiang, G., DeGrip, W. J., Hayes, W. P. & Rollag, M. D. Melanopsin: an opsin in melanophores, brain, and eye. *Proc. Natl Acad. Sci. USA* **95**, 340–345 (1998).
- Provincio, I. *et al.* A novel human opsin in the inner retina. *J. Neurosci.* **20**, 600–605 (2000).
- Gooley, J. J., Lu, J., Chou, T. C., Scammell, T. E. & Saper, C. B. Melanopsin in cells of origin of the retinohypothalamic tract. *Nature Neurosci.* **4**, 1165 (2001).
- Hannibal, J., Hindersson, P., Knudsen, S. M., Georg, B. & Fahrenkrug, J. The photopigment melanopsin is exclusively present in pituitary adenylate cyclase-activating polypeptide-containing retinal ganglion cells of the retinohypothalamic tract. *J. Neurosci.* **22**, RC191 (2002).
- Provincio, I., Rollag, M. D. & Castrucci, A. M. Photoreceptive net in the mammalian retina. *Nature* **415**, 493 (2002).
- Hattar, S., Liao, H.-W., Takao, M., Berson, D. M. & Yau, K.-W. Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science* **295**, 1065–1070 (2002).
- Berson, D. M., Dunn, F. A. & Takao, M. Phototransduction by retinal ganglion cells that set the circadian clock. *Science* **295**, 1070–1073 (2002).
- Lucas, R. J. *et al.* Diminished pupillary light reflex at high irradiances in melanopsin-knockout mice. *Science* **299**, 245–247 (2003).
- Panda, S. *et al.* Melanopsin (*Opn4*) requirement for normal light-induced circadian phase shifting. *Science* **298**, 2213–2216 (2002).
- Ruby, N. F. *et al.* Role of melanopsin in circadian responses to light. *Science* **298**, 2211–2213 (2002).
- Lucas, R. J., Douglas, R. H. & Foster, R. G. Characterization of an ocular photopigment capable of driving pupillary constriction in mice. *Nature Neurosci.* **4**, 621–626 (2001).
- Emery, P., So, W. V., Kaneko, M., Hall, J. C. & Roshash, M. CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell* **95**, 669–679 (1998).
- Ceriani, M. F. *et al.* Light-dependent sequestration of TIMELESS by CRYPTOCHROME. *Science* **285**, 553–556 (1999).
- Selby, C. P., Thompson, C., Schmitz, T. M., van Gelder, R. N. & Sancar, A. Functional redundancy of cryptochromes and classical photoreceptors for nonvisual ocular photoreception in mice. *Proc. Natl Acad. Sci. USA* **97**, 14697–14702 (2000).
- van Gelder, R. N. Non-visual ocular photoreception. *Ophthalmic Genet.* **22**, 195–205 (2001).
- van Gelder, R. N., Wee, R., Lee, J. A. & Tu, D. C. Reduced pupillary light responses in mice lacking cryptochromes. *Science* **299**, 222 (2003).
- Griffin, E. A. Jr, Staknis, D. & Weitz, C. J. Light-independent role of CRY1 and CRY2 in the mammalian circadian clock. *Science* **286**, 768–771 (1999).
- Lucas, R. J., Freedman, M. S., Munoz, M., Garcia-Fernandez, J. M. & Foster, R. Regulation of the mammalian pineal by non-rod, non-cone, ocular photoreceptors. *Science* **284**, 505–507 (1999).
- Calvert, P. D. *et al.* Phototransduction in transgenic mice after targeted deletion of the rod transducin α-subunit. *Proc. Natl Acad. Sci. USA* **97**, 13913–13918 (2000).
- Biel, M. *et al.* Selective loss of cone function in mice lacking the cyclic nucleotide-gated channel CNG3. *Proc. Natl Acad. Sci. USA* **96**, 7553–7557 (1999).
- Bradley, J., Frings, S., Yau, K.-W. & Reed, R. Nomenclature for ion channel subunits. *Science* **294**, 2095–2096 (2001).
- Cone, R. A. Early receptor potential: photoreversible charge displacement in rhodopsin. *Science* **155**, 1128–1131 (1967).
- Murakami, M. & Pak, W. L. Intracellularly recorded early receptor potential of the vertebrate photoreceptors. *Vision Res.* **10**, 965–975 (1970).

24. Brockerhoff, S. E. *et al.* Light stimulates a transducin-independent increase of cytoplasmic Ca^{2+} and suppression of current in cones from the zebrafish mutant *nof*. *J. Neurosci.* **23**, 470–480 (2003).
25. Wee, R., Castrucci, A. M., Provencio, I., Gan, L. & van Gelder, R. N. Loss of photic entrainment and altered free-running circadian rhythms in *math5^{-/-}* mice. *J. Neurosci.* **22**, 10427–10433 (2002).
26. Aschoff, J. in *Trends in Chronobiology. Advances in the Biosciences* Vol. 73 (eds Hekken, W. T. J. M., Kerkhof, G. A. & Rietveld, W. J.) 149–161 (Pergamon, Oxford, 1988).
27. Mrosovsky, N. Masking: history, definitions, and measurement. *Chronobiol. Int.* **16**, 415–429 (1999).
28. Mrosovsky, N., Lucas, R. J. & Foster, R. G. Persistence of masking responses to light in mice lacking rods and cones. *J. Biol. Rhythms* **16**, 585–587 (2001).
29. Redlin, U. & Mrosovsky, N. Masking of locomotor activity in hamsters. *J. Comp. Physiol. A* **184**, 429–437 (1999).
30. Mrosovsky, N. Further characterization of the phenotype of *mCry1/mCry2*-deficient mice. *Chronobiol. Int.* **18**, 613–625 (2001).
31. Yoshimura, T. & Ebihara, S. Spectral sensitivity of photoreceptors mediating phase-shifts of circadian rhythms in retinally degenerate CBA/J (*rd/rd*) and normal CBA/N (+/+) mice. *J. Comp. Physiol. A* **178**, 797–802 (1996).
32. Provencio, I. & Foster, R. G. Circadian rhythms in mice can be regulated by photoreceptors with cone-like characteristics. *Brain Res.* **694**, 183–190 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H.-W. Liao and J. Lai for help in performing the immunocytochemical experiments with antibodies against cryptochromes; J. Butler for help with real-time RT-PCR; Y. Liang for help in genotyping the animals; P. Salmon for help in the wheel-running experiments; and D. M. Berson, H. R. Matthews, G. L. Fain and members of the Yau laboratory for scientific discussions. This work was supported by the US National Eye Institute (K.-W.Y.), the UK Biotechnology and Biological Sciences Research Council and the Wellcome Trust (R.J.L. and R.G.F.), and the Canadian Institutes of Health Research (N.M.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.-W.Y. (kwya@u@mail.jhmi.edu).

A TRPV family ion channel required for hearing in *Drosophila*

Janghwan Kim[†], Yun Doo Chung[‡], Dae-young Park^{*}, Sookyoung Choi^{*}, Dong Wook Shin^{*}, Heun Soh^{||}, Hye Won Lee^{*}, Wonseok Son^{*}, Jeongbin Yim^{||}, Chul-Seung Park^{||}, Maurice J. Kern[‡] & Changsoo Kim^{*}

^{*} Department of Genetics, Hanwha Chemical Co. R&D Center, Sinsung-Dong, Yusong-Gu, Daejeon 305-345, Korea

[†] Department of Microbiology, Chungnam National University, Daejeon 305-764, Korea

[‡] Department of Neurobiology & Behavior, and [§] Center for Developmental Genetics, State University of New York at Stony Brook, New York 11794, USA

^{||} Department of Life Science, K-JIST, Gwangju 500-712, Korea

[¶] National Creative Research Institute for Genetic Reprogramming, Seoul National University, Seoul 151-742, Korea

The many types of insect ear share a common sensory element, the chordotonal organ, in which sound-induced antennal or tympanal vibrations are transmitted to ciliated sensory neurons and transduced to receptor potentials^{1,2}. However, the molecular identity of the transducing ion channels in chordotonal neurons, or in any auditory system, is still unknown^{3,4}. *Drosophila* that are mutant for NOMPC, a transient receptor potential (TRP) superfamily ion channel, lack receptor potentials and currents in tactile bristles^{5,6} but retain most of the antennal sound-evoked response⁷, suggesting that a different channel is the primary transducer in chordotonal organs. Here we describe the *Drosophila* Nanchung (Nan) protein, an ion channel subunit similar to vanilloid-receptor-related (TRPV) channels of the TRP superfamily. Nan mediates hypo-osmotically activated calcium influx and cation currents in cultured cells. It is expressed *in vivo* exclusively in chordotonal neurons and is localized to their

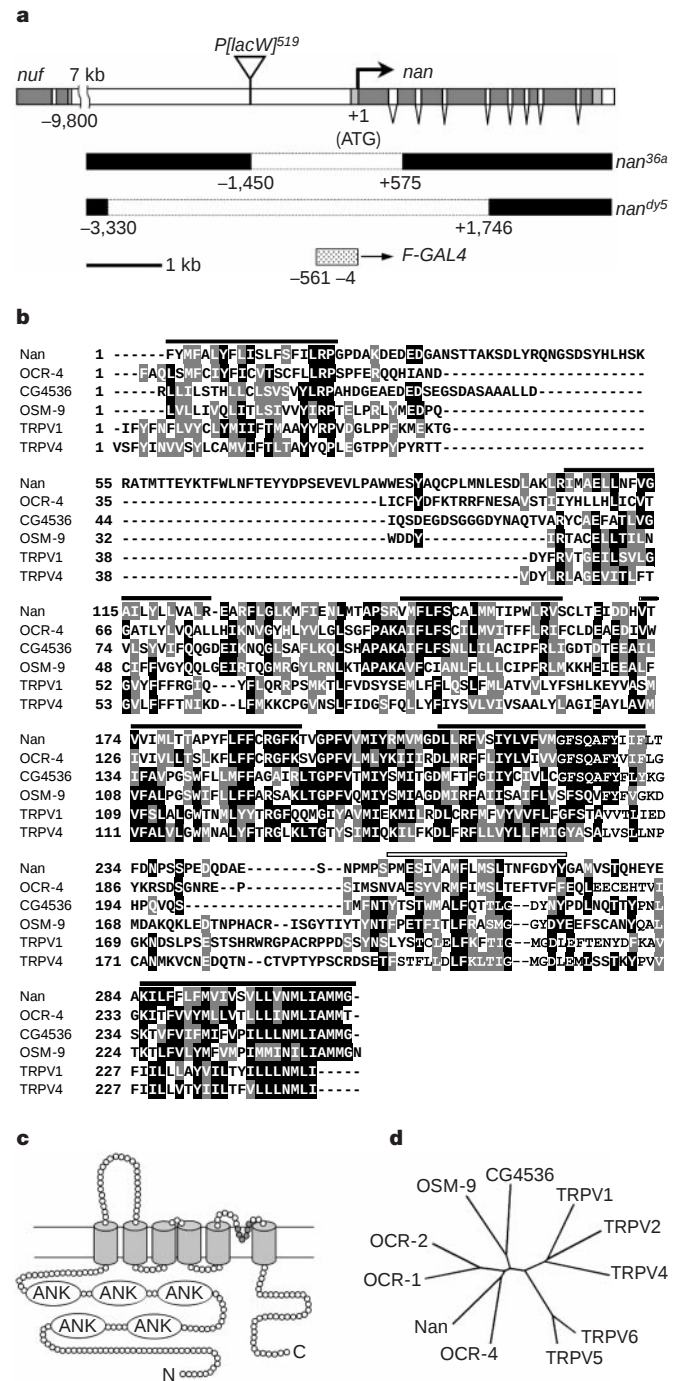


Figure 1 Molecular analysis of *nan*. **a**, Genomic organization. A transcription start site at -91 with respect to the translation start (+1) is indicated (angled arrow). A P transposon (*P[lacW]⁵¹⁹*) inserted 1.4-kb upstream of the transcription start site was imprecisely excised to give the deletions *nan^{36a}* and *nan⁴⁵*, deleting 150 and 478 codons, respectively. The *F-GAL4* construct contains a 557-bp fragment (dotted box).

b, Alignment of the transmembrane region from TRPV proteins, including *Drosophila* Nan and CG4536, *C. elegans* OCR-4 and OSM-9, and human TRPV1 and TRPV4. Transmembrane domains (closed bars) and a putative pore region (open bar) were predicted by TMHMM2.0 (ref. 25). Black shading, identical amino acids; grey shading, similar amino acids; as used by Boxshade program (www.ch.embnet.org). **c**, Topology model for Nan protein, based on SMART²⁶ and TMHMM2.0 predictions. The putative pore region is denoted by a closed circle. **d**, Phylogenetic tree of the TRPV ion channel subfamily, based on a CLUSTALW alignment.

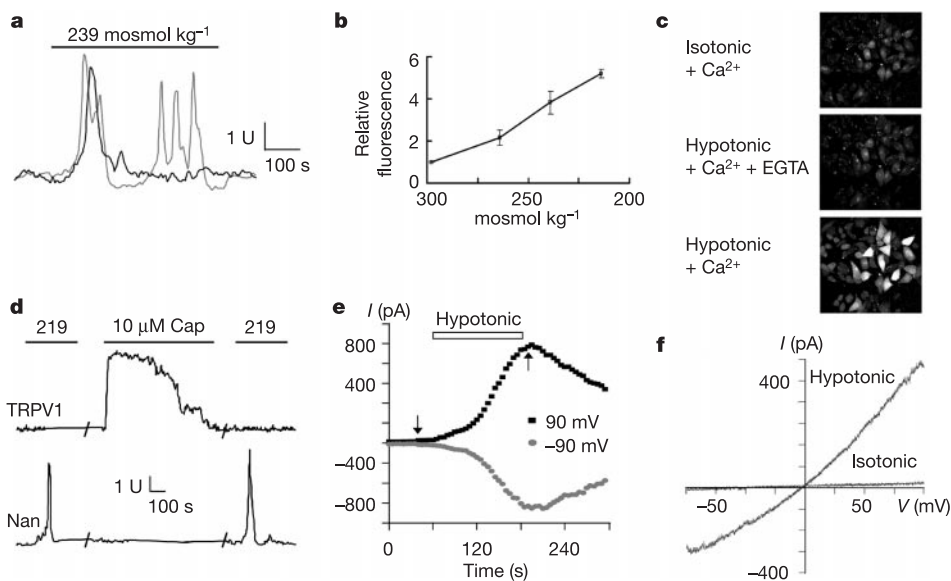


Figure 2 Nan mediates cation influx in response to hypo-osmotic stress. **a**, Nan-expressing cells show single Ca^{2+} peaks (black) or oscillations (grey) during hypo-osmotic stimulation ($239 \text{ mosmol kg}^{-1}$, 2 mM Ca^{2+}). U, arbitrary unit of Ca^{2+} signals measured by confocal microscopic analysis. **b**, Intracellular Ca^{2+} increases as osmolality decreases. The mean amplitudes of the first Ca^{2+} peak in response to hypotonic solutions (219 , 239 , $264 \text{ mosmol kg}^{-1}$) are normalized by the Ca^{2+} signals in isotonic solution ($298 \text{ mosmol kg}^{-1}$). Error bars represent s.e.m. Data were recorded from five independent

experiments. **c**, Addition of EGTA (10 mM) to the 2 mM Ca^{2+} hypotonic medium eliminated the Ca^{2+} increase. Replacement of the EGTA-containing solution with Ca^{2+} hypotonic solution restored the Ca^{2+} signal. **d**, Nan-expressing cells are unresponsive to capsaicin. TRPV1-expressing cells respond to capsaicin ($10 \mu\text{M}$) but not hypotonic solution. **e**, **f**, Ionic currents and I - V relationship stimulated by a hypotonic solution ($221 \text{ mosmol kg}^{-1}$) in Nan-expressing cells in a whole-cell patch configuration.

sensory cilia. Antennal sound-evoked potentials are completely absent in mutants lacking Nan, showing that it is an essential component of the chordotonal mechanotransducer.

Channels of the TRPV subfamily may be activated by physical stimuli including osmotic stress and heat, as well as by vanilloid, phorbol and lipid ligands^{8,9}. They have been implicated in sensory transduction in both nematodes and mammals. The first TRPV protein identified, *Caenorhabditis elegans* OSM-9, is expressed in ciliated neurons and is required for their transduction of diverse sensory stimuli including nose touch and aversive hyperosmotic solutions, as well as for olfactory adaptation¹⁰. Co-expression of OSM-9 and another TRPV subunit, OCR-2, is required for their localization in cilia and function in transduction¹¹. In mammals, the heat-activated capsaicin receptor TRPV1 and the related TRPV2 are expressed in different subsets of sensory neurons and are involved in non-mechanical nociception^{12–14}, whereas the osmotically activatable channel TRPV4 is expressed in inner-ear hair cells, Merkel cells in whiskers, sensory neurons and brain osmosensory regions, in addition to ion-transporting epithelia¹⁵.

The sequenced *Drosophila melanogaster* genome¹⁶ includes two predicted TRPV genes, CG4536 and CG5842 (refs 11, 17). Complementary DNA from CG5842, which encodes Nan, was sequenced (Fig. 1) and found to encode a predicted protein of 833 amino acids with the characteristic TRPV domain architecture: five cytoplasmic ankyrin repeats precede six putative transmembrane domains and a pore region (Fig. 1c). The extracellular loop between the first and second predicted transmembrane domain is about 60 amino acids longer than in other TRPV proteins. Nan is most similar in sequence to *C. elegans* OCR-4 (37% amino acid identity) with decreasing similarity to the other OCR proteins, *Drosophila* CG4536 and *Caenorhabditis* OSM-9, and the mammalian TRPV channels (Fig. 1d).

To test whether Nan can form an ion channel, Chinese hamster ovary (CHO-K1) cells were transfected with *nan* cDNA and loaded with the calcium indicator fluo-4/AM. Transfected cells, but not

mock-transfected controls, showed an increase in intracellular calcium after superfusion with a hypo-osmotic solution (Fig. 2a). The response, which appeared in different individual cells as single peaks or oscillations (Fig. 2a), increased with decreasing osmolality (Fig. 2b) and was dependent on extracellular calcium (Fig. 2c). As previously observed^{15,18}, similarly treated cells expressing the human vanilloid receptor TRPV1 showed rises in intracellular Ca^{2+} in response to capsaicin (Fig. 2d) and high temperature (40 – 60°C , data not shown), but not to hypo-osmotic bath solution (Fig. 2d). Conversely, Nan-expressing cells were unresponsive to capsaicin (Fig. 2d) and high temperature (data not shown). Hypo-osmotically-induced currents were recorded from transfected cells in whole-cell patch configuration (Fig. 2e, f). With symmetrical sodium as the charge carrier, a slightly outward-rectifying current was induced by hypo-osmotic solution increasing in amplitude over a few minutes of superfusion. No such currents were observed in mock-transfected cells (data not shown). Under bi-ionic conditions, the ratio of K^+ to Na^+ permeability ($P_{\text{K}}/P_{\text{Na}}$) was 0.84 ± 0.09 (standard error of the mean (s.e.m.); data not shown). These data indicate that Nan forms a cation channel that can be activated by hypotonic stress.

nan expression in *Drosophila* embryos was examined by *in situ* hybridization, revealing transcripts only in chordotonal organs (Fig. 3a). To visualize the *nan*-expressing cells, a 557-base pair (bp) fragment including the *nan* promoter was fused to DNA encoding the yeast transcription activator Gal4, and this construct (*F-GAL4*, Fig. 1a) was used to drive expression of green fluorescent protein (GFP) from *UAS-eGFP* transgenes. Transgenic embryos carrying both constructs showed GFP expression in chordotonal neurons after late stage 16. No expression was seen in any other cells of the peripheral nervous system. Expression was similarly specific to chordotonal neurons in adults, including those in the femur (Fig. 3c), antenna (Fig. 3d) and wing (data not shown).

To determine the subcellular location of Nan in chordotonal

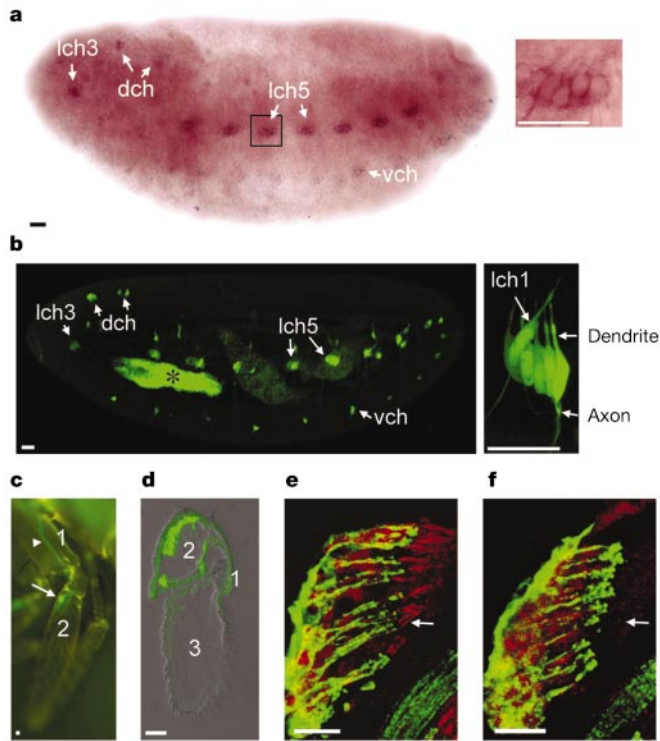


Figure 3 Expression of Nan in chordotonal organs. **a**, *In situ* hybridization of a *nan* antisense RNA probe to embryos at late stage 16. Dorsal (dch), lateral (lch) and ventral (vch) chordotonal organs are indicated. The boxed region is enlarged in the inset. **b–d**, Images of *F-GAL4/UAS-eGFP* flies. **b**, Confocal image of embryo at late stage 16. Dorsal, lateral and ventral chordotonal organs are indicated. A single lch cluster is shown in the inset. Staining in the salivary gland (asterisk) is due to the vector used for *F-GAL4* (ref. 27). **c**, Fluorescence image of an adult leg. The coxa (1) and femur (2) are numbered. The arrow indicates GFP-expressing neurons of the femoral chordotonal organ; the arrowhead indicates a labelled axonal projection. **d**, Combined fluorescence and bright-field images of a sectioned antenna showing strong GFP expression in the chordotonal organ. **e, f**, Cryostat sections of the second antennal segment, labelled with anti-Nan serum (red) and 22C10 antibody (green). Ciliary outer segments (arrow) of chordotonal neurons are stained in wild type (**e**) but not in *nan* mutants (**f**). Scale bars, 10 μ m.

neurons, polyclonal antisera were raised against two overlapping amino-terminal segments (amino acids 1–130 and 53–244). Both antisera specifically stained the ciliary outer segments of chordotonal neurons (Fig. 3e). No staining was observed in *nan* mutants (Fig. 3f).

Mutations in *nan* were generated by imprecise excision of a P transposon inserted upstream of the gene. Two deletions extending into *nan*, *nan*^{36a} and *nan*^{dy5}, were recovered (Fig. 1a). Each deleted the transcription start site and extended into the open reading frame. Homozygotes and trans-heterozygotes for these mutations are viable as adults, but show abnormal behaviour: they are sedentary (Fig. 4a), mildly uncoordinated, more likely than wild-type flies to fall when climbing, and slower to right themselves after a fall (observations not shown). This behavioural phenotype is similar to that of mutations specifically affecting chordotonal organ differentiation⁷, suggesting a defect in chordotonal organ function.

To test auditory transduction, extracellular sound-evoked potentials were recorded from the antennal chordotonal organ (Johnston's organ). This organ comprises many individual sensory units or scolopidia, which are stimulated when distal antennal segments are vibrated by near-field sound⁷. Trains of sound pulses elicited corresponding response potentials in *nan*/+ heterozygotes, but no

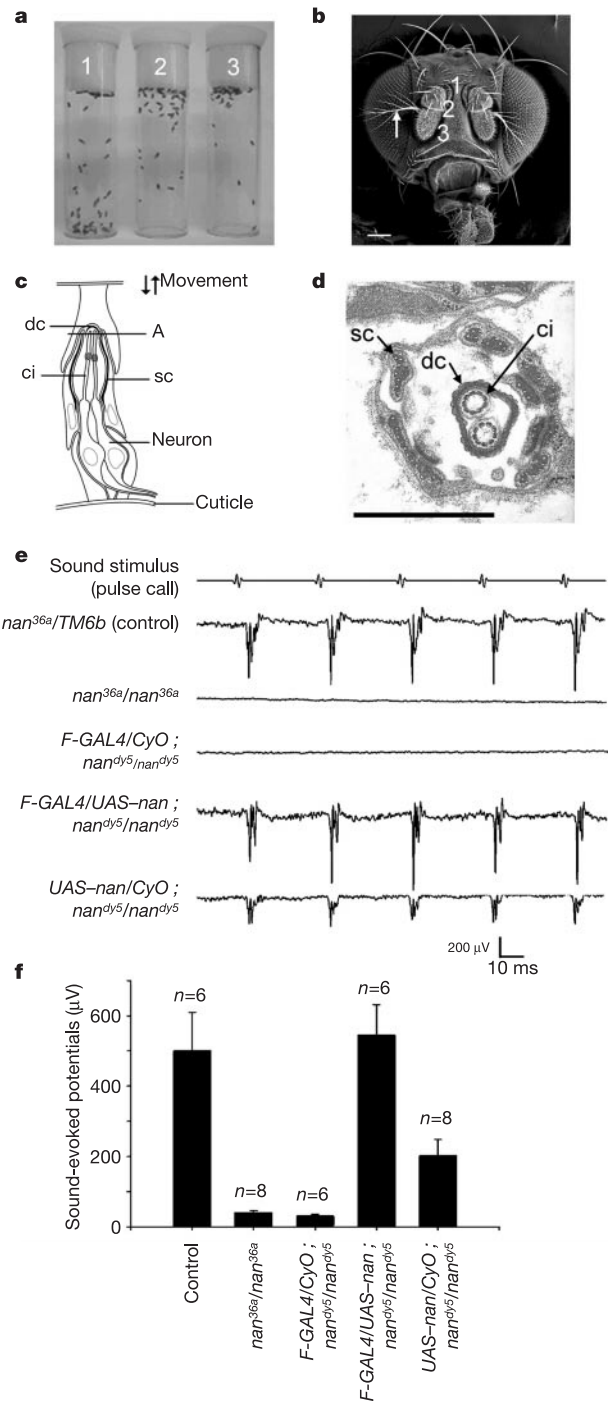


Figure 4 Behavioural and electrophysiological defects in *nan* mutants. **a**, Sedentary behaviour of *nan* mutants. Flies were tapped to the bottom of the tube and observed after 1 min. Tube 1, *nan*^{36a}/*nan*^{dy5}; tube 2, *nan*^{36a}/*nuf*¹; tube 3, *nuf*¹/*nuf*¹. *nuf* is the neighbouring gene to *nan*. **b**, Scanning electron microscopic image of an adult head showing the antennal segments (numbered) and the arista (arrow). Scale bar, 100 μ m. **c**, Schematic drawing of an antennal chordotonal organ. The scolopales (sc), dendritic caps (dc) and sensory cilia (ci) are indicated. **d**, Ultrathin transverse section of a *nan* mutant antenna as indicated by line A in **c**. Scale bar, 2 μ m. **e**, Antennal extracellular potentials recorded during sound stimulation. Each trace is an average of the responses to ten consecutive stimulus trains of five pulses. Normal sound-evoked potentials in heterozygote controls are absent in *nan* homozygotes, and are restored by *F-GAL4* driving *UAS-nan*. Two independent inserts of *UAS-nan* show partial rescue in the absence of *F-GAL4*. **f**, Histogram of maximum peak amplitudes from averaged traces. *n*, number of antennae.

response was observed in *nan* mutant homozygotes (Fig. 4e), even when stimulus intensity was increased and many trials were averaged (data not shown). The sound-evoked potentials were fully restored in mutants by a *UAS-nan* cDNA transgene when driven by *F-GAL4* (Fig. 4e, f). Notably, this transgene also restored partial function in the absence of *F-GAL4* (Fig. 4e). This may be due to the presence within the 5' untranslated region of two copies of a putative binding site for the transcription factor RFX, which regulates the differentiation of ciliated sensory neurons^{19,20}.

To check whether the absence of sound-evoked potentials was due to a defect in specification or differentiation of chordotonal neurons or support cells, mutant and wild-type antennae were examined by scanning and transmission electron microscopy. No structural or ultrastructural defects were seen in mutant chordotonal organs. In particular, the scolopidial elements required to transmit and transduce mechanical stimuli, including the neuronal cilium, scolopale and dendritic cap^{7,21}, all appeared normal (Fig. 4d).

The location of Nan in mechanosensory cilia, and the combination in *nan* mutants of defective transduction with normal sense organ morphology, strongly suggest that Nan forms part of a chordotonal transducer channel. Similarly to the mammalian TRPV4 channel, it combines expression in mechanosensory cells with activation by cellular hypotonic stress. Activation of Nan by hypo-osmotic stress may in fact reflect sensitivity to membrane tension, with mechanically induced ciliary membrane stretch also gating the channel *in situ*. Alternatively, osmotic activation and mechanotransduction may represent different modes of channel activation. The TRPV4-mediated response to hypo-osmotic solutions may depend on phosphorylation²²; however, it is unlikely that such a pathway could be modulated fast enough to achieve the sub-millisecond latencies characteristic of sensory mechanotransducers. TRPV channels are typically activatable by multiple stimulus modes: TRPV4 is activated by heat²³ and by phorbol ligands²⁴ in addition to hypotonic stress. For Nan, as for all putative mechanosensory channels identified so far, a direct correspondence between a specific mode of channel activation and the induction of mechanoreceptor currents *in situ* remains to be established. □

Methods

Molecular biology

A transcription start site was determined by 5' rapid amplification of cloned ends (Ambion). *UAS-nan* contains a 2.5-kilobase (kb) cDNA from upstream (−257) to the termination site, which was prepared from Canton-S poly-A-tailed RNA using Superscript II (Invitrogen) and amplified by polymerase chain reaction (PCR) (Turbo Pfu, Stratagene). *F-GAL4* contains sequence (557 bp) of the upstream region amplified by PCR and ligated into pCasPer-Gal4.

Generation of anti-Nan sera

Amino acids 1–130 and 53–244 of Nan were fused to the carboxy terminus of glutathione S-transferase (GST) and maltose-binding protein (MBP), respectively. Fused proteins were expressed in *Escherichia coli* and injected into rats.

Measurement of intracellular Ca²⁺

nan full-length cDNA was subcloned into pcDNA3.1 mycis(−) (Invitrogen). CHO-K1 cells were transfected with the *nan*-encoding vector, and stable cell lines were selected for resistance to G418 (500 ng ml^{−1}) (Invitrogen). Intracellular Ca²⁺ concentration was measured as described¹⁵.

Patch-clamp recordings

Whole-cell currents were recorded from CHO-K1 cells with an Axopatch 200B amplifier (Axon Instruments). Patch recordings were performed at room temperature 24–48 h after transfection. Borosilicate glass patch pipettes had resistances of 3–5 MΩ when filled with the isotonic intracellular solution. To measure channel currents, cells were held at a potential of 0 mV and subjected to 1-s voltage ramps from −100 mV to 100 mV. Signals were filtered at 1 kHz using a four-pole low-pass Bessel filter, digitized at 200 samples per ms using Digidata 1200 (Axon Instruments), and stored in a personal computer. pClamp8 software (Axon Instruments) was used to control the amplifier and to acquire the data. Pipette and bath solutions contained 5 mM HEPES, 1 mM EDTA, 120 mM NaCl and 100 mM mannitol. To activate the currents, cells were

superfused while recording with a hypotonic solution containing 5 mM HEPES, 1 mM EDTA and 120 mM NaCl. In bi-ionic conditions, 120 mM KCl was substituted for NaCl in the bath solutions.

Antennal recordings

Extracellular sound-evoked potentials were recorded with sharpened tungsten electrodes placed in the antenna and head, amplified and digitized as previously described⁷. Mutant flies were recorded in alternation with wild-type controls. Sound stimuli, which approximate the pulse phase of *Drosophila* courtship calls, were generated and delivered under near-field conditions, as described⁷.

Received 9 February; accepted 8 May 2003; doi:10.1038/nature01733.

Published online 18 June 2003.

1. Yager, D. D. Structure, development, and evolution of insect auditory systems. *Microsc. Res. Tech.* **47**, 380–400 (1999).
2. Eberl, D. F. Feeling the vibes: chordotonal mechanisms in insect hearing. *Curr. Opin. Neurobiol.* **9**, 389–393 (1999).
3. Caldwell, J. C. & Eberl, D. F. Towards a molecular understanding of *Drosophila* hearing. *J. Neurobiol.* **53**, 172–189 (2002).
4. Strassmaier, M. & Gillespie, P. G. The hair cell's transduction channel. *Curr. Opin. Neurobiol.* **12**, 380–386 (2002).
5. Kernan, M. J., Cowan, D. & Zuker, C. Genetic dissection of mechanosensory transduction: mechanoreception-defective mutations of *Drosophila*. *Neuron* **12**, 1195–1206 (1994).
6. Walker, R., Willingham, A. & Zuker, C. A *Drosophila* mechanosensory transduction channel. *Science* **287**, 2229–2234 (2000).
7. Eberl, D., Hardy, R. & Kernan, M. J. Genetically similar transduction mechanisms for touch and hearing in *Drosophila*. *J. Neurosci.* **20**, 5981–5988 (2000).
8. Gunthorpe, M. J., Benham, C. D., Randall, A. & Davis, J. B. The diversity in the vanilloid (TRPV) receptor family of ion channels. *Trends Pharmacol. Sci.* **23**, 183–191 (2002).
9. Benham, C. D., Davis, J. B. & Randall, A. D. Vanilloid and TRP channels: a family of lipid-gated cation channels. *Neuropharmacology* **42**, 873–888 (2002).
10. Colbert, H., Smith, T. & Bargmann, C. OSM-9, a novel protein with structural similarity to channels, is required for olfaction, mechanosensation, and olfactory adaptation in *Caenorhabditis elegans*. *J. Neurosci.* **17**, 8259–8369 (1997).
11. Tobin, D. *et al.* Combinatorial expression of TRPV channel proteins defines their sensory functions and subcellular localization in *C. elegans* neurons. *Neuron* **35**, 307–318 (2002).
12. Caterina, M. *et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816–824 (1997).
13. Caterina, M. J. *et al.* Impaired nociception and pain sensation in mice lacking the capsaicin receptor. *Science* **288**, 306–313 (2000).
14. Caterina, M. J., Rosen, T. A., Tominaga, M., Brake, A. J. & Julius, D. A capsaicin-receptor homologue with a high threshold for noxious heat. *Nature* **398**, 436–441 (1999).
15. Liedtke, W. *et al.* Vanilloid receptor-related osmotically activated channel (VR-OAC), a candidate vertebrate osmoreceptor. *Cell* **103**, 525–535 (2000).
16. Adams, M. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
17. Littleton, J. T. & Ganetzky, B. Ion channels and synaptic organization: analysis of the *Drosophila* genome. *Neuron* **26**, 35–43 (2000).
18. Strotmann, R., Harteneck, C., Nunnenmacher, K., Schultz, G. & Plant, T. D. OTRPC4, a nonselective cation channel that confers sensitivity to extracellular osmolarity. *Nature Cell Biol.* **2**, 695–702 (2000).
19. Vandaele, C., Coulon-Bublex, M., Couble, P. & Durand, B. *Drosophila* regulatory factor X is an embryonic type I sensory neuron marker also expressed in spermatids and in the brain of *Drosophila*. *Mech. Dev.* **103**, 159–162 (2001).
20. Dubruille, R. *et al.* *Drosophila* regulatory factor X is necessary for ciliated sensory neuron differentiation. *Development* **129**, 5487–5498 (2002).
21. Chung, Y., Zhu, J., Han, Y. & Kernan, M. *nompA* encodes a PNS-specific, ZP-domain protein required to connect mechanosensory dendrites to sensory structures. *Neuron* **29**, 415–428 (2001).
22. Xu, H. *et al.* Regulation of a TRP channel by tyrosine phosphorylation: Src family kinase-dependent phosphorylation of TRPV4 on Y253 mediates its response to hypotonic stress. *J. Biol. Chem.* **278**, 11520–11527 (2003).
23. Guler, A. D. *et al.* Heat-evoked activation of the ion channel, TRPV4. *J. Neurosci.* **22**, 6408–6414 (2002).
24. Watanabe, H. *et al.* Activation of TRPV4 channels (hVR1-2/mTRP12) by phorbol derivatives. *J. Biol. Chem.* **277**, 13569–13577 (2002).
25. Moller, S., Croning, M. D. & Apweiler, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646–653 (2001).
26. Schultz, J., Milpetz, F., Bork, P. & Ponting, C. P. SMART, a simple modular architecture research tool: Identification of signalling domains. *Proc. Natl Acad. Sci. USA* **95**, 5857–5864 (1998).
27. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).

Acknowledgements We thank K.-W. Choi and S. Lee for comments on a preliminary version of this manuscript; W. Sullivan for *nuf*¹; and the Korea Basic Science Institute for use of equipment for Ca²⁺ measurement and electron microscope. This work was supported in part by a National Institute of Deafness and Communicative Disorders grant to M.K., by a Korea Research Foundation Grant to C.S.P. and by a Creative Research Initiatives of the Korean Ministry of Science and Technology grant to J.Y.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.K. (changgk2001@hanmail.net). The *nan* sequence is deposited in GenBank under accession number AY262004.

Assessing the redundancy of MADS-box genes during carpel and ovule development

Anusak Pinyopich*, Gary S. Ditta*, Beth Savidge*, Sarah J. Liljegren*, Elvira Baumann†, Ellen Wisman† & Martin F. Yanofsky*

* Division of Biological Sciences, University of California at San Diego, La Jolla, California 92093-0116, USA

† Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné-Weg 10, 50829 Köln, Germany

Carpels are essential for sexual plant reproduction because they house the ovules and subsequently develop into fruits that protect, nourish and ultimately disperse the seeds. The *AGAMOUS* (*AG*) gene is necessary for plant sexual reproduction because stamens and carpels are absent from *ag* mutant flowers^{1,2}. However, the fact that sepals are converted into carpelloid organs in certain mutant backgrounds even in the absence of *AG* activity indicates that an *AG*-independent carpel-development pathway exists². *AG* is a member of a monophyletic clade of MADS-box genes that includes *SHATTERPROOF1* (*SHP1*), *SHP2* and *SEEDSTICK* (*STK*)³, indicating that these four genes might share partly redundant activities. Here we show that the *SHP* genes are responsible for *AG*-independent carpel development. We also show that the *STK* gene is required for normal development of the funiculus, an umbilical-cord-like structure that connects the developing seed to the fruit, and for dispersal of the seeds when the fruit matures. We further show that all four members of the *AG* clade are required for specifying the identity of ovules, the landmark invention during the course of vascular plant evolution that enabled seed plants to become the most successful group of land plants⁴.

The fact that *ag* mutants completely lack carpels demonstrates the primary role of *AG* in specifying carpel identity^{1,2}. Mutations in the *APETALA2* (*AP2*) gene, which encodes a transcription factor for a putative protein of the same type as ethylene-responsive element-binding proteins, lead to the ectopic expression of *AG* throughout the flower⁵ and the replacement of sepals (Fig. 1a) by organs that closely resemble carpels (Fig. 1b). The observation that ectopic carpelloid organs develop in place of sepals in *ag ap2* double mutants indicates that many features of carpels can form in the absence of *AG* activity⁶. These features include the formation of stigmatic papillae, style, replum and placental tissues with ovules (Fig. 1c). Because *ag* single mutants lack carpels, we used *ap2 ag* double mutants to identify the genes that might be responsible for *ag*-independent carpel development.

The carpel- and fruit-specific *SHP1* and *SHP2* genes are good candidates for genes that act redundantly with *AG* because they share extensive sequence similarity and partly overlapping expression patterns with *AG*. In addition, the *SHP* genes are expressed in the ectopic carpels of *ap2* mutants^{7,8}. To determine whether the *SHP* genes are involved in the formation of ectopic carpelloid organs in *ag ap2* mutants, we constructed *ap2 ag shp1 shp2* quadruple mutants. Remarkably, all features of carpels, including ovules, were completely absent from the first-whorl organs of the quadruple mutants (Fig. 1d, e). Although a small number of outgrowths along the margins of these organs were observed in the quadruple mutants, these outgrowths had no ovule-like characteristics and contained cuticular ridges and guard cells that are never observed on ovules (Fig. 1f). These data demonstrate that the *SHP* genes are responsible for *AG*-independent carpel and ovule development observed in *ap2 ag* mutants. Although the carpel-specific expression of the *SHP* genes is normally dependent on *AG* activity⁷,

these data also indicate that the activation of the *SHP* genes in the first-whorl organs of *ap2* mutants is independent of *AG*. Previous studies indicated that the SPATULA (*SPT*) basic helix-loop-helix transcription factor has a function in *AG*-independent carpel

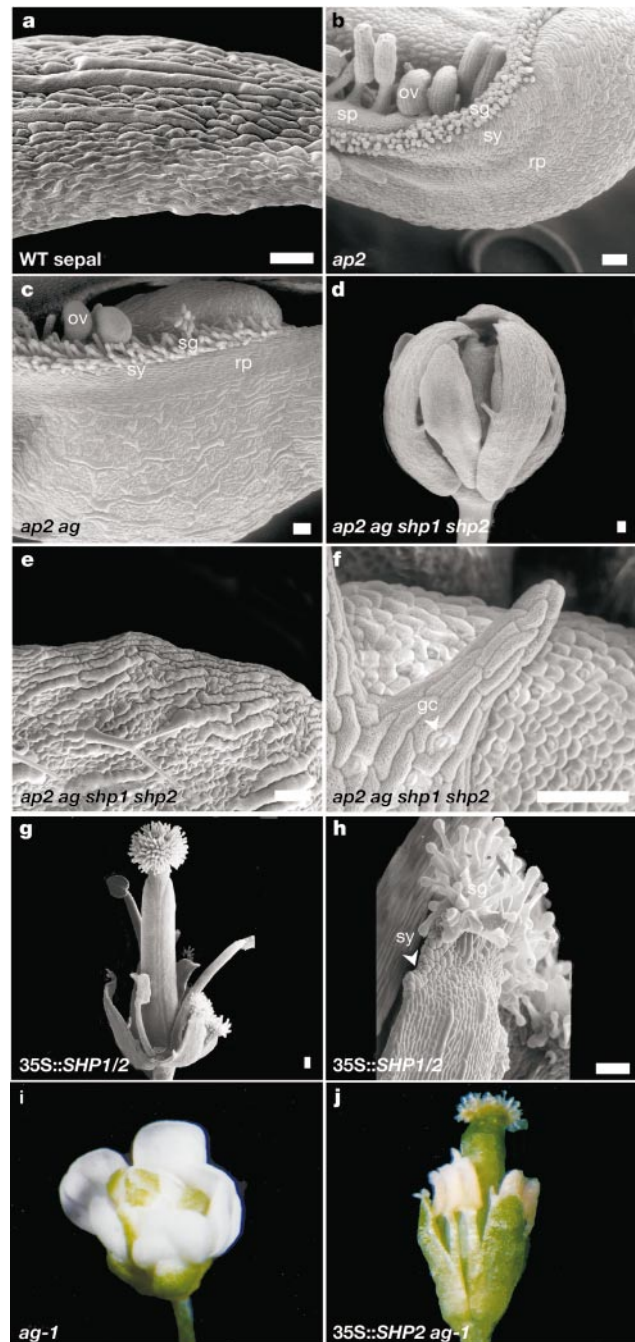


Figure 1 *AG* acts redundantly with *SHP* genes to specify carpel identity. **a**, Margin of wild-type (WT) sepal. **b**, **c** Several carpel cell types: stigma (sg), style (sy), ovule (ov), septum (sp) and replum (rp) are observed at the margin of first-whorl organ of *ap2* (**b**) and *ap2 ag* mutants (**c**). **d–f**, First-whorl organs of *ap2 ag shp1 shp2* mutants lack all carpelloid features and have a few small outgrowths with cuticular ridges and guard cells (gc) on the epidermal surface (**f**) that are never observed on wild-type ovules. The *ap2 ag shp1 shp2* mutant is indistinguishable from the *ap2 ag shp1 shp2 stk* mutant (data not shown). **g**, **h**, Constitutive expression of *SHP1/2* genes causes the conversion of sepals toward carpels. Stigmatic tissue (sg) and style-like cells (sy) were observed at the margin of first-whorl organs. **i**, **j**, Overexpression of *SHP2* can rescue stamen and carpel development (**j**) in *ag* mutants (**i**). The rescued carpel produced viable seeds when pollinated with wild-type pollen (see Supplementary Information). Scale bars, 50 μ m.

development because mutations in *SPT* largely eliminate the formation of ectopic carpelloid organs⁶. These genetic data, together with the observation that *SPT* expression at the margin of carpelloid organs in *ap2* mutants is reduced in the *ap2 ag* double-mutant background⁹, support the suggestion that *AG* and *SHP* might redundantly regulate *SPT* expression.

To determine whether the *SHP* genes are sufficient for promoting carpel development in ectopic positions, we generated transgenic plants in which the *SHP* genes were 'constitutively' expressed from the cauliflower mosaic virus 35S promoter¹⁰. 35S::*SHP1/2* plants showed a partial conversion of first-whorl sepals toward carpels, as demonstrated by the extensive proliferation of stigmatic papillae as well as cells resembling those normally found on the style of wild-type carpels (Fig. 1g, h). These and other phenotypes, such as the transformation of petals toward stamens (Fig. 1g), early flowering, curly leaves and prematurely open flower buds, are similar to those previously described for 35S::*AG* plants¹¹, indicating that ectopic expression of the *SHP* genes is sufficient to confer many of the same phenotypes caused by ectopic *AG* expression (see Supplementary Information).

What is the underlying mechanism that led to the evolution of partly overlapping, yet distinct, functions for the *SHP* and *AG* genes? One possibility is that these genes, which are the apparent result of relatively recent gene duplication events, have diverged in their expression patterns even though the proteins themselves have retained similar activities. To test this idea, we introduced the 35S::*SHP2* transgene into *ag* mutants. Remarkably, constitutive expression of the *SHP2* gene was sufficient to largely rescue stamen and carpel development in *ag* mutants (Fig. 1i, j). The fact that the *SHP2* gene can substitute for *AG* in promoting stamen development is particularly striking because *SHP2* is normally not expressed in stamens. These data demonstrate that the *SHP* genes have retained the ability to substitute for *AG* activity and that the differences

between the activities of the *SHP* and *AG* genes can largely be attributed to their different patterns of expression.

In addition to their roles in promoting *AG*-independent carpel and ovule development, the *SHP* genes are required for the differentiation of the dehiscence zone in mature fruit¹⁰. Separation of the dehiscence-zone cells is required for the normal seed-dispersal process. The *SHP* genes, which are expressed in the dehiscence zones of developing fruit, are also expressed in ovules^{7,8}. However, *shp1 shp2* mutant ovules seem normal, indicating that a redundant factor might hide their possible roles during ovule development. One candidate for a redundant factor is the *STK* gene (formerly known as *AGL11*; ref. 12), because it shares extensive sequence similarity with the *SHP* genes¹³ and because its expression domain overlaps that of the *SHP* genes during ovule development^{7,8,12}.

Initially, *STK* RNA is uniformly expressed in placental tissues and ovule primordia¹². *STK::GUS* expression mirrors the pattern of *STK* RNA accumulation (Fig. 2a–c) and GUS activity is strongest in the funiculus, a stalk-like structure that connects the ovule and seed to the remainder of the fruit (Fig. 2d). In mature ovules, *STK::GUS* is expressed strongly in the funiculus and weakly in integuments that will later form the seed coat (Fig. 2e). This expression pattern continues after fertilization until stage 15, when GUS activity is no longer detectable.

To uncover the possible roles of the *STK* gene during ovule and seed development, we first identified *stk* loss-of-function mutants (see Methods). One of the most striking phenotypes observed in *stk* single mutants is a failure of the seeds to be released from the mature fruits (Fig. 2g). Seed detachment normally occurs by separation of the seed body from the funiculus at a site referred to as the seed abscission zone. In wild-type fruit, the seed abscission zone is located immediately adjacent to the seed body and seems constricted owing to the smaller size of cells in this region than in

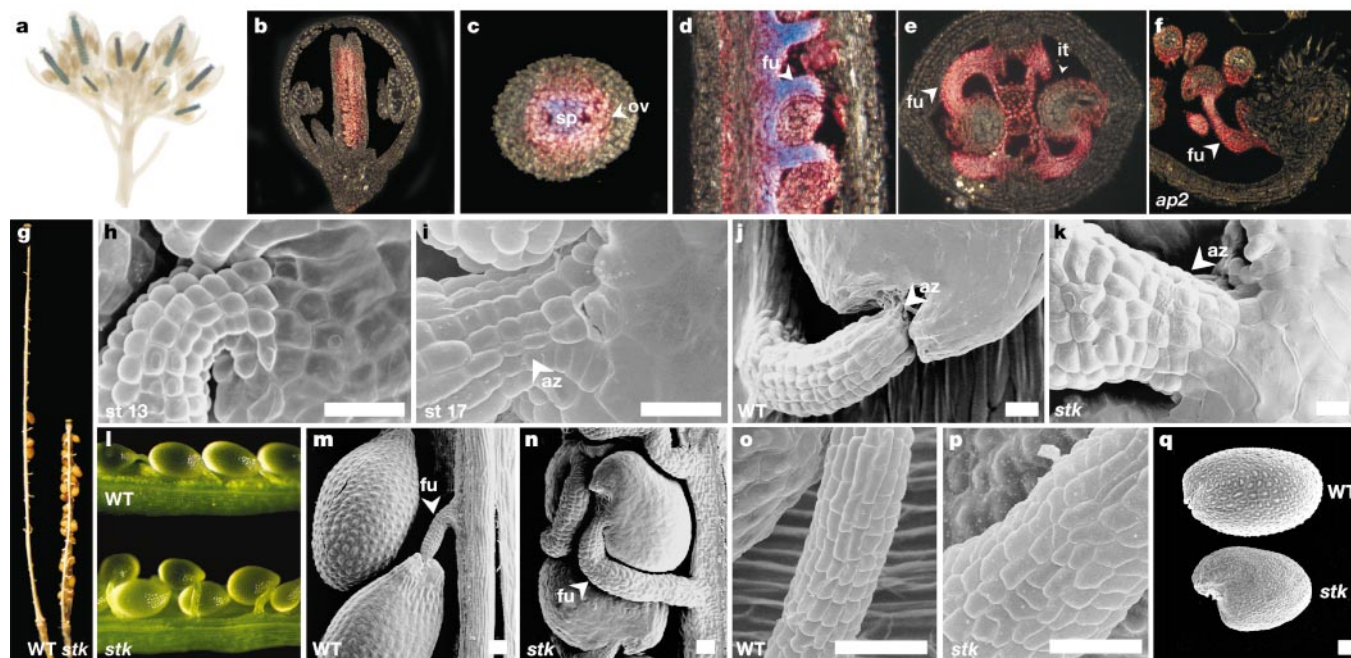


Figure 2 Characterization of *STK*. **a**, Staining of inflorescence for GUS reveals strong expression in carpels but not in other organs. **b**, GUS activity is first detected in placental tissue and young ovule primordia within carpels (stage 8). **c**, Transverse section of a carpel, showing uniform GUS expression in ovule primordia (ov) and septum (sp) (stage 9). **d**, During stage 11–12, expression is strong (blue) in the funiculus (fu) and weaker in the rest of the ovule (pink). **e**, At stage 13, GUS staining is restricted to funiculus and integuments (it). **f**, GUS activity is present in ectopic ovules on first-whorl organs of *ap2*

mutants. **g**, Wild-type (WT) fruit (left) and *stk* fruit at stage 19–20. Most seeds stay attached to the fruit in *stk* mutants. **h**, **i**, Differentiation of seed abscission zone (az) is observed at stage 17 (**i**) but not at stage 13 (**h**). **j**, **k**, Seed abscission (**j**) fails to occur in *stk* mutants (**k**). **l**, Arrangement of the developing seeds is irregular in *stk* mutants. **m**, **n**, At stage 17, wild-type funiculus (**m**) is much smaller than *stk* mutant funiculus (**n**). **o**, **p**, Differences in funicular cell size of wild-type cells (**o**) and mutant cells (**p**). **q**, Mature *stk* mutant seeds are smaller than wild-type seeds. Scale bars, 20 μ m (**h–k**); 50 μ m (**m–q**).

neighbouring tissue (Fig. 2i). Differentiation of abscission-zone cells occurs after fertilization, because no constriction is observed at the same region of mature wild-type ovules (Fig. 2h). Proper differentiation of these cells seems to be disrupted in *stk* mutants, because the separation of abscission zone cells fails to occur (Fig. 2j, k).

The *stk* mutant fruits are shorter than those of the wild type, typically reaching only 60% of the wild-type length. Inside the developing fruit, the regular positioning of seeds that occurs in the wild type is perturbed in *stk* mutants (Fig. 2l). This irregular seed spacing is apparently caused by a drastically enlarged funiculus in the *stk* mutants (Fig. 2m, n). Close inspection by scanning electron microscopy revealed that mutant funicular cells are larger than in the wild type (Fig. 2o, p) and that the cell number is increased. These results show that *STK* is required to prevent the abnormal growth of the funiculus by controlling cell expansion and cell division. In addition, whereas wild-type seeds adopt an oblong shape, *stk* seeds are rounder and smaller (Fig. 2q).

To determine whether *STK* acts redundantly with the *SHP* genes to control ovule development, we constructed *stk shp1 shp2* triple mutants. Remarkably, normal ovule and seed development was completely disrupted in these mutants. Although some viable but abnormal seeds were obtained, many of the ovules and seeds were arrested during their development. The most striking phenotype is that some of the ovules were converted into either leaf-like or carpel-like structures (Fig. 3a, b). Close inspection of cells on the surface of these converted ovules revealed style-like and valve-like cells characteristic of carpels that are never observed on wild-type

ovules (Fig. 3c, d). Occasionally, stigmatic tissue was also observed among the converted ovules. Conversion of ovules into carpeloid structures has also been reported in *Petunia* when the activities of the two putative *STK* gene orthologues *Floral Binding Protein7* (*FBP7*) and *FBP11* were removed simultaneously¹⁴. Moreover, ectopic expression of *FBP11* in *petunia*¹⁵ as well as *STK* in *Arabidopsis* (L. Colombo, personal communication) was sufficient to induce the formation of ectopic ovules. Taken together, these results demonstrate that the *STK* gene is both necessary and sufficient to promote ovule development.

Because the *STK* (Fig. 2f), *SHP* and *AG* genes are all expressed in the ectopic ovules that form in the first-whorl organs of *ap2* mutants^{5,7,8}, we investigated their roles in promoting ectopic ovule development. Although many of the ectopic ovules that form in *ap2* mutants develop normally, about 40% of them develop into carpeloid structures¹⁶. More ovules (55%) are converted in *ap2 ag* double mutants, confirming previously published results of the role for *AG* in promoting ovule identity¹⁶. The conversion of ovules into carpeloid structures occurs even more frequently (81%) in *ap2 stk ag* triple mutants (Fig. 3e), indicating that *STK* and *AG* might have redundant roles in promoting ovule identity. Loss of ovule identity is strongest in *ap2 stk shp1 shp2* quadruple mutants (Fig. 3f), because nearly all (95%) ectopic ovules were converted into carpeloid structures. Although the remaining ovules have integument-like characteristics, they did not develop into mature ovules. Taken together, these data demonstrate that all four members of the *AG* clade contribute to proper ovule development.

These studies have revealed the redundant roles of *AG* and the *SHP* genes in promoting carpel development, as well as the redundant roles of *AG*, *STK* and the *SHP* genes in promoting ovule identity (Fig. 4). Although these four genes clearly evolved from a common ancestral gene and retain redundant activities, they have now evolved specific functions, with *AG* promoting reproductive organ identity, the *SHP* genes promoting fruit dehiscence, and the *STK* gene controlling funiculus development and seed detachment. It is interesting that both fruit dehiscence and seed abscission, which are two distinct steps in seed dispersal process, are controlled by closely related MADS-box genes.

The occurrence of large multigene families has emerged as a common theme now that the genome sequences of *Arabidopsis* and other organisms are known. This fact has highlighted the genetic redundancy that probably occurs in closely related family members. The MADS-box gene family in *Arabidopsis* serves as a good example of this redundancy, because it contains dozens of members, many of

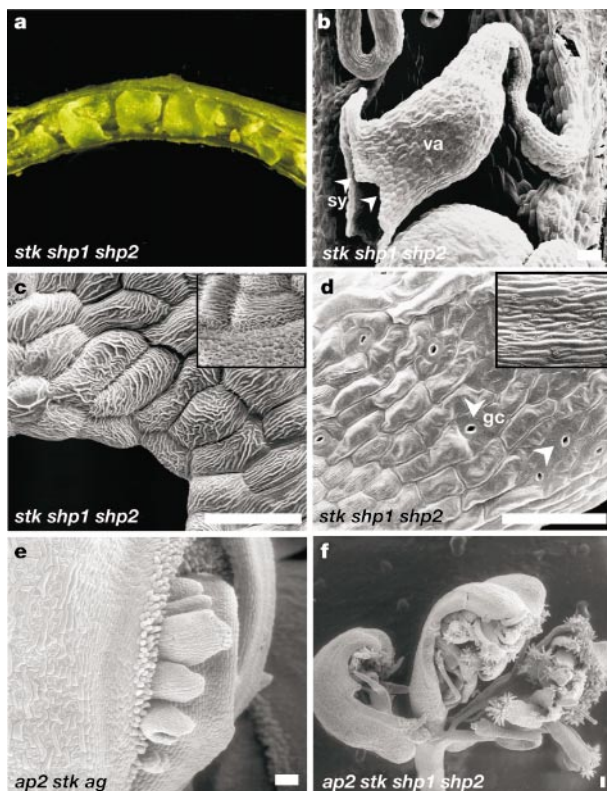


Figure 3 All members of the *AG* clade redundantly specify ovule identity. **a, b**, Stage-17 *stk shp1 shp2* mutant fruit reveals conversion of many ovules into carpel-like structures (**b**) that have style-like (sy) regions located at the edges and valve-like (va) regions in the middle. **c**, Close-up of the tip of the carpel-like structure in (**b**) showing style-like cells with cuticular thickening, a characteristic of wild-type stilar cells (inset). **d**, Cells in the middle part of the carpel-like structure in (**b**) resemble valve cells of wild-type carpel (inset). Several guard cells (gc) are indicated. **e**, Most of the ectopic ovules are converted into carpel-like structures in *ap2 stk ag* mutants. **f**, Ectopic ovules are largely absent from first-whorl organs of *ap2 stk shp1 shp2* mutants. Scale bars, 50 μ m (except **c**, 20 μ m).

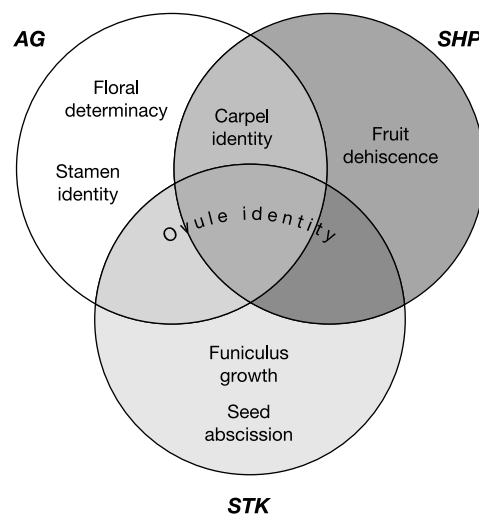


Figure 4 Unique and overlapping functions of genes in *AGAMOUS* clade.

which are known to control aspects of plant development. The AG MADS-box gene has a critical function in plant sexual reproduction, because *ag* mutants completely lack reproductive organs. However, because AG is only one of four members of a monophyletic clade, it has been necessary to characterize all possible combinations of single, double, triple and quadruple mutants. Now that the individual and redundant roles of the AG clade of MADS-box genes are known, we can begin to explore their roles in setting up the patterns of downstream gene expression required for carpel and ovule development. □

Methods

Reporter construct and histology

The STK::GUS reporter construct is a translational fusion of sequence 2 kilobases (kb) upstream of the STK translation start codon with the GUS (*uidA*) gene in pDW137 (ref. 17). The sequence was amplified by primer P1 (5'-CCTGCAGCGTGATTGCAATTGCAAAATTGAG-3') and P5 (5'-CTCTAGACCCATCTTCAATTTAAACA-3') then cloned into pCRII vector, yielding pPOP163. This 2-kb sequence contains 1.3 kb of the first intron and 0.7 kb of the promoter region. The sequence was then cloned into the *Pst*I and *Xba*I sites of pDW137, resulting in pPOP166. GUS staining was performed as described previously¹⁸ except that the concentrations of potassium ferrocyanide and potassium ferricyanide were 5 mM.

Plant materials

The *stk-1* allele was identified by screening for an *En-1* insertion among a collection of plants carrying about 50,000 independent insertions by using the STK-specific primer 11-X5 (5'-CCACTAACCATTGTGATGATGGTGTGT-3') and the *En-1*-specific primer En8130 (5'-GAGCGTCGGTCCCACTTCTATAC-3')¹⁹. *En-1* is inserted near the 3' end of exon 3 (2 base pairs (bp) from the splice site). The *stk-2* and *stk-3* stable alleles were obtained by screening for germinal excision of the *stk-1* allele. The *stk-2* allele contains a 74-bp insertion near the splice site of third intron. The *stk-3* allele has a 47-bp deletion that removes the splice site. Analysis by polymerase chain reaction (PCR) with reverse transcription showed that mutations in both alleles affected RNA splicing (data not shown). Two additional mutant alleles, *stk-4* and *stk-5*, were obtained from the Madison Alpha collection by using the gene-specific primers 11-X15 (5'-TTCTTGCAGTCTGCCACTAGTTTGTGTGT-3') and 11-X18 (5'-AACTGCTTCGTTACGCACCGAACTCAACA-3'), respectively, together with the T-DNA-specific primer JL202 (5'-CATTTTATAATAACGCTGCGGACATCTAC-3'). The *stk shp1 shp2* triple mutants were generated by crossing *shp1-1 shp2-1* double-mutant plants with plants carrying either the *stk-2* or the *stk-3* allele. The *shp* alleles^{10,20} *ap2-6* (ref. 21) and *ag-2* (ref. 22) used for generating the mutants were described previously. The 35S:*SHP1/2* transgenic plants were described previously¹⁰. The 35S:*SHP2* transgene was introduced into *ag-1* mutants.

Mutant genotyping

The *stk-2* and *stk-3* alleles were genotyped by PCR with the primers 11-x4 (5'-GCTTGTCTGATAGCACCACACTAGCA-3') and 11-x5 (5'-CCACTAACCATTGTGATGATGGTGTGT-3'). The *stk shp1 shp2* triple mutants were identified by screening for mutant phenotypes and later confirmed by PCR genotyping all three genes. Genotyping of *shp1-1* and *shp2-1* allele was as described previously¹⁰. The *ap2-6* allele was genotyped by PCR amplification with the primers G1021 (5'-GCAGCAGCTCGGTATTTTTC-3') and G1022 (5'-ATCAACCGGTGGTTCTCAG-3') followed by digestion with *Hind*III. The *Hind*III site is present only in the *ap2-6* allele.

Scanning electron microscopy

Flowers and fruits were fixed overnight in 1.6% osmium tetroxide in 50 mM sodium cacodylate buffer at 4 °C. Tissues were washed twice in the same buffer and dehydrated in an acetone series. After critical-point drying, pistils and fruits were dissected to expose ovules or seeds inside. Tissues were coated with gold/palladium. Samples were examined in either a Cambridge S360 scanning electron microscope or a Quanta 600 microscope using an acceleration voltage of 5–20 kV.

Received 26 February; accepted 1 May 2003; doi:10.1038/nature01741.

1. Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. Genes directing flower development in *Arabidopsis*. *Plant Cell* **1**, 37–52 (1989).
2. Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development* **112**, 1–20 (1991).
3. Theissen, G. et al. A short history of MADS-box genes in plants. *Plant Mol. Biol.* **42**, 115–149 (2000).
4. Stewart, W. N. & Rothwell, G. W. *Paleobotany and the evolution of plants* (Cambridge Univ. Press, 1993).
5. Drews, G. N., Bowman, J. L. & Meyerowitz, E. M. Negative regulation of the *Arabidopsis* homeotic gene *AGAMOUS* by the *APETALA2* product. *Cell* **65**, 991–1002 (1991).
6. Alvarez, J. & Smyth, D. R. *CRABS CLAW* and *SPATULA*, two *Arabidopsis* genes that control carpel development in parallel with *AGAMOUS*. *Development* **126**, 2377–2386 (1999).
7. Savidge, B., Rounsley, S. D. & Yanofsky, M. F. Temporal relationship between the transcription of two *Arabidopsis* MADS box genes and the floral organ identity genes. *Plant Cell* **7**, 721–733 (1995).
8. Flanagan, C. A., Hu, Y. & Ma, H. Specific expression of the *AGL1* MADS-box gene suggests regulatory functions in *Arabidopsis* gynoecium and ovule development. *Plant J.* **10**, 343–353 (1996).
9. Heisler, M. G. B., Atkinson, A., Bylstra, Y. H., Walsh, R. & Smyth, D. R. *SPATULA*, a gene that controls

- development of carpel margin tissues in *Arabidopsis*, encodes a bHLH protein. *Development* **128**, 1089–1098 (2001).
10. Liljegren, S. J. et al. *SHATTERPROOF* MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* **404**, 766–770 (2000).
11. Mizukami, Y. & Ma, H. Ectopic expression of the floral homeotic gene *AGAMOUS* in transgenic *Arabidopsis* plants alters floral organ identity. *Cell* **71**, 119–131 (1992).
12. Rounsley, S. D., Ditta, G. S. & Yanofsky, M. F. Diverse roles for MADS box genes in *Arabidopsis* development. *Plant Cell* **7**, 1259–1269 (1995).
13. Ma, H., Yanofsky, M. F. & Meyerowitz, E. M. *AGL1-AGL6*, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* **5**, 484–495 (1991).
14. Angenent, G. C. et al. A novel class of MADS box genes is involved in ovule development in *Petunia*. *Plant Cell* **7**, 1569–1582 (1995).
15. Colombo, L. et al. The *Petunia* MADS box gene *FBP11* determines ovule identity. *Plant Cell* **7**, 1859–1868 (1995).
16. Western, T. L. & Haughn, G. W. *BELL1* and *AGAMOUS* genes promote ovule identity in *Arabidopsis thaliana*. *Plant J.* **18**, 329–336 (1999).
17. Blazquez, M. A., Soowal, L. N., Lee, I. & Weigel, D. *LEAFY* expression and flower initiation in *Arabidopsis*. *Development* **124**, 3835–3844 (1997).
18. Sessions, A., Weigel, D. & Yanofsky, M. F. The *Arabidopsis thaliana* *MERISTEM LAYER 1* promoter specifies epidermal expression in meristems and young primordia. *Plant J.* **20**, 259–263 (1999).
19. Wisman, E., Cardon, G. H., Fransz, P. & Saedler, H. The behaviour of the autonomous maize transposable element *En/Spm* in *Arabidopsis thaliana* allows efficient mutagenesis. *Plant Mol. Biol.* **37**, 989–999 (1998).
20. Kempin, S. A. et al. Targeted disruption in *Arabidopsis*. *Nature* **389**, 802–803 (1997).
21. Kunst, L., Klenz, J. E., Martinezpater, J. & Haughn, G. W. *AP2* gene determines the identity of perianth organs in flowers of *Arabidopsis thaliana*. *Plant Cell* **1**, 1195–1208 (1989).
22. Yanofsky, M. F. et al. The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. *Nature* **346**, 35–39 (1990).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Ray for advice on scanning electron microscopy, D. Weigel for providing the pDW137 vector, members of Yanofsky laboratory for comments, and H. Chang and P. Golshani for technical assistance. A.P. received a scholarship from the Ananda Mahidol Foundation, and this work was supported by a grant from the National Science Foundation (to M.Y.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.Y. (marty@ucsd.edu).

Selective imprinting of gut-homing T cells by Peyer's patch dendritic cells

J. Rodrigo Mora*, Maria Rosa Bono†, N. Manjunath*, Wolfgang Weninger*, Lois L. Cavanagh*, Mario Roseblatt† & Ulrich H. von Andrian*

* The Center for Blood Research and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

† Laboratorio de Inmunología, Facultad de Ciencias, Universidad de Chile, Fundación Ciencia para la Vida, and Millennium Institute for Fundamental and Applied Biology (MIFAB), Santiago 6842301, Chile

Whereas naive T cells migrate only to secondary lymphoid organs^{1,2}, activation by antigen confers to T cells the ability to home to non-lymphoid sites^{3,4}. Activated effector/memory T cells migrate preferentially to tissues that are connected to the secondary lymphoid organs where antigen was first encountered^{5–7}. Thus, oral antigens induce effector/memory cells that express essential receptors for intestinal homing, namely the integrin $\alpha 4\beta 7$ and CCR9, the receptor for the gut-associated chemokine TECK/CCL25 (refs 6, 8, 9). Here we show that this imprinting of gut tropism is mediated by dendritic cells from Peyer's patches. Stimulation of CD8-expressing T cells by dendritic cells from Peyer's patches, peripheral lymph nodes and spleen induced equivalent activation markers and effector activity in T cells, but only Peyer's patch dendritic cells induced high levels of $\alpha 4\beta 7$, responsiveness to TECK and the ability to home to the small

intestine. These findings establish that Peyer's patch dendritic cells imprint gut-homing specificity on T cells, and thus license effector/memory cells to access anatomical sites most likely to contain their cognate antigen.

It has been suggested that, during antigen stimulation of naive lymphocytes, the lymphoid tissue microenvironment provides instructions for peripheral homing preference^{1,5,6}. However, this concept has been questioned¹⁰. We reasoned that, if antigen presentation has a lymphoid-organ-specific 'flavour', tissue tropism might be conferred to naive T cells by professional antigen-presenting cells, particularly dendritic cells (DCs). Thus, we analysed CD8-expressing (CD8⁺) T cells from P14 × T-GFP mice that were stimulated *in vitro* by antigen-pulsed DCs from different lymphoid tissues. The P14 transgenic T-cell receptor (TCR) recognizes the gp33–41 peptide from lymphocytic choriomeningitis virus (LCMV) in H-2D^b (major histocompatibility complex (MHC) class I). In P14 × T-GFP mice, TCR-transgenic naive T cells express green fluorescent protein (GFP), whereas differentiated cytotoxic T lymphocytes (CTLs) downregulate GFP, providing a marker to distinguish effector cells from non-effector subsets⁴.

To expand DCs in lymphoid tissues, donor mice were implanted subcutaneously with melanoma cells that secrete Flt3-L (FMS-like tyrosine kinase 3 ligand)¹¹. Two weeks later, DCs were isolated from lymphoid organs by immunomagnetic depletion of other haematopoietic lineages. DCs were pulsed with gp33–41 peptide and co-cultured with naive CD8⁺ P14 × T-GFP cells. One week later, CD8⁺ cells that had been stimulated by DCs from Peyer's patches (PP), peripheral lymph nodes (PLN) and spleen (henceforth referred to as CD8^{PP-DC}, CD8^{PLN-DC} and CD8^{Sp-DC}, respectively) had more than doubled in number, expressed low levels of GFP (GFP^{low}) or were GFP-negative (GFP[−]), expressed several common activation markers and displayed the 1B11 epitope of CD43, which distinguishes effector CTLs from naive and resting memory CD8⁺ cells (Table 1 and Fig. 1). Indeed, CD8^{PP-DC}, CD8^{PLN-DC} and, to a slightly lesser degree, CD8^{Sp-DC} exerted antigen-specific cytotoxicity and produced interleukin-2 (IL-2) and interferon-γ on restimulation. Thus, all three DC populations generated bona fide effector CTLs.

Despite these functional and phenotypic similarities, there was a striking difference in α4β7 integrin expression. Although all stimulated T-cell populations expressed higher levels of α4β7 than did naive T cells, cells that expressed high levels of α4β7 (α4β7^{high} cells) were much more frequent among CD8^{PP-DC} cells than among CD8^{PLN-DC} or CD8^{Sp-DC}. This difference was apparent whether DCs were enriched by negative selection (85–90% CD11c⁺) or purified by FACS sorting (>98% CD11c⁺; Supplementary Fig. S-1a, b). By contrast, α4β7^{high} T cells were not increased when the sorted

CD11c[−] fraction from enriched PP DCs was added to CD8⁺ cells during stimulation with anti-CD3/CD28 (Supplementary Fig. S-1c) or with PLN DCs or spleen DCs (not shown). Thus, the induction of α4β7^{high} CTLs was mediated by DCs and not by other components of PP.

This unique effect of PP DCs is significant because only α4β7^{high} effector cells can attach efficiently to MadCAM-1 (mucosal vascular addressin cell-adhesion molecule 1), which is constitutively expressed at relatively low density in intestinal microvessels^{12,13}. Indeed, α4β7 levels comparable to those on CD8^{PP-DC} cells were found on subsets of TCRαβ⁺CD8αβ⁺ memory cells in murine spleen and small intestine (not shown). However, it is unlikely that peptide-pulsed PP DCs expanded selectively the rare α4β7^{high} memory cells within the P14 × T-GFP starting population, because induction of α4β7 was equivalent when PP DCs were co-cultured with α4β7[−] naive CD8⁺ P14 × T-GFP cells that were sorted to >99% purity (Supplementary Fig. S-2). Moreover, 90% of α4β7^{high} CD8^{PP-DC} cells expressed the P14 TCR, as assessed by staining with gp33–41-H-2D^b MHC class I tetramer (Supplementary Fig. S-3). Also, PP DCs induced high levels of α4β7 on ovalbumin-specific T cells from OT-I × Rag^{−/−} mice (Supplementary Fig. S-5a), in which conventional effector/memory cells are absent¹⁴. Thus, PP DCs did not merely expand memory cells that may have been preprogrammed for homing to the gut, but induced intestinal homing receptors during a primary T-cell response independent of TCR specificity.

Week-long exposure of T cells to DCs from any lymphoid tissue induced higher α4β7 levels than were observed on naive T cells. This is consistent with a recent report that T-cell activation with anti-CD3 enhanced α4β7 levels, and expression increased further in the presence of mesenteric lymph node (MLN) DCs¹⁵. On the other hand, whereas α4β7^{high} CTLs arose *in vitro* only after >4 days (not shown), immunization *in vivo* upregulates α4β7 on CD4⁺ cells within 2 days, and only in intestinal lymphoid tissues, whereas CD4⁺ cells in cutaneous lymph node downregulate α4β7 (ref. 7). This suggests that, in intact lymphoid organs, additional factors that are not fully recapitulated in our system might modulate the kinetics and tissue specificity of α4β7 induction.

Nevertheless, the difference in α4β7 staining between CD8^{PP-DC} and other CTL populations was highly specific, as it was most apparent with the monoclonal antibody (MAb) DATK32, which recognizes the α4β7 heterodimer¹⁶ (Table 1). Differences in the individual α4 and β7 chains were less pronounced, presumably because these subunits can also dimerize with integrins β1 and αE, respectively¹⁷. Indeed, β1 integrins were only modestly higher on CD8^{PP-DC}, whereas αE expression was similar on all CTL

Table 1 Surface phenotype of naive and activated P14 × T-GFP cells

	CD8 ^{PP-DC}	CD8 ^{PLN-DC}	CD8 ^{Sp-DC}	CD8 ^{naive}
Cell number (×10 ⁶ ml ^{−1})	2.2 ± 0.6	2.7 ± 0.8	2.7 ± 0.9	NA
Gp33–41-H-2D ^b tetramer binding (%)	88 ± 3	92 ± 1	87 ± 5	80 ± 13
TCR-Vα2 chain (%)	168 ± 49 (93 ± 2)	177 ± 45 (94 ± 1)	128 ± 10 (94 ± 1)	190 ± 8 (92 ± 5)
CD25	174 ± 76	180 ± 81	122 ± 90	3 ± 2***
CD44	996 ± 367	891 ± 342	962 ± 211	115 ± 16***
CD122	18 ± 4	19 ± 6	14 ± 3	4 ± 1***
CD69 (%)	25 ± 10 (17 ± 8)	34 ± 15 (23 ± 14)	35 ± 13 (32 ± 14)	5 ± 1*** (3 ± 1)***
1B11/CD43	363 ± 192	393 ± 249	447 ± 279	70 ± 26***
L-selectin (%)	51 ± 33 (26 ± 9)	78 ± 51 (30 ± 10)	128 ± 98** (39 ± 8)**	319 ± 103*** (95 ± 4)***
GFP	31 ± 9	41 ± 17	54 ± 26	316 ± 82***
LFA-1	861 ± 321	970 ± 381	1,024 ± 396	319 ± 98***
β1 integrin	177 ± 17	142 ± 16*	134 ± 13*	14 ± 10***
αE integrin (%)	284 ± 108 (82 ± 11)	282 ± 113 (74 ± 13)	238 ± 61 (70 ± 13)	72 ± 68*** (43 ± 28)***
α4β7 heterodimer (% α4β7 ^{high})	165 ± 68 (52 ± 12)	65 ± 21*** (16 ± 8)***	49 ± 25*** (9 ± 6)***	15 ± 8*** (0.3 ± 0.2)***
α4 integrin	137 ± 93	71 ± 41**	69 ± 43**	31 ± 18***
β7 integrin	598 ± 184	468 ± 141	382 ± 106*	139 ± 90***
PSGL-1	69 ± 17	55 ± 18	60 ± 18	18 ± 5***

Results for gp33–41-H-2D^b tetramer staining are shown as percentage of binding cells. Mean fluorescence intensity (MFI) of markers for T-cell activation (CD25, CD44, CD122, CD45RB, CD69), effector differentiation (GFP, 1B11) and adhesion molecules were measured on freshly isolated naive T cells and DC-stimulated effector cells (day 6–7) after gating on CD8α⁺ T cells in 10–22 independent experiments (mean ± s.d.). When two distinct populations were present, the percentage of positive cells is shown in parentheses. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 compared with CD8^{PP-DC}. NA, not applicable; PSGL-1, P-selectin glycoprotein ligand-1.

populations. Likewise, the $\beta 2$ integrin LFA-1 (lymphocyte-function-associated antigen) was upregulated equally in all CTL samples, whereas L-selectin, a homing receptor for PLN, was uniformly downregulated.

Next, we asked whether other gut-specific traffic molecules are preferentially modulated by PP DCs. In particular, CCR9, the receptor for TECK, has been implicated in T-cell migration to the small intestine^{9,18,19}. Because antibodies to murine CCR9 were unavailable, we compared CTL migration towards a TECK gradient (Fig. 2a). $CD8^{PLN-DC}$ and $CD8^{Sp-DC}$ migrated poorly to TECK, whereas $CD8^{PP-DC}$ responded well to this chemokine. This effect was consistently induced by immunomagnetically enriched and FACS-purified PP DCs (Supplementary Fig. S-4), and was also observed with OT-I $\times$ RAG^{-/-} T cells (Supplementary Fig. S-5b).

By contrast, all CTL populations responded equally to several other lymphoid and inflammatory chemokines (Fig. 2b).

As shown previously, naive $CD8^+$ cells also migrated towards TECK²⁰. Their response was even greater than that of $CD8^{PP-DC}$. This suggests that DCs from PLN and spleen induced a loss of CTL responsiveness to TECK that was partially prevented or reversed by PP DCs. Accordingly, naive T cells expressed more CCR9 messenger RNA than did CTLs (Fig. 2c). Interestingly, CCR9 mRNA levels were lower in $CD8^{PLN-DC}$ than in $CD8^{PP-DC}$, but they were equivalent in $CD8^{PP-DC}$ and $CD8^{Sp-DC}$. Thus, differential transcriptional regulation of CCR9 explains only partially the observed differences in chemotactic efficacy of TECK. This implies a role for post-transcriptional regulation of CCR9 expression or function, analogous to that reported previously for CCR6 (ref. 21).

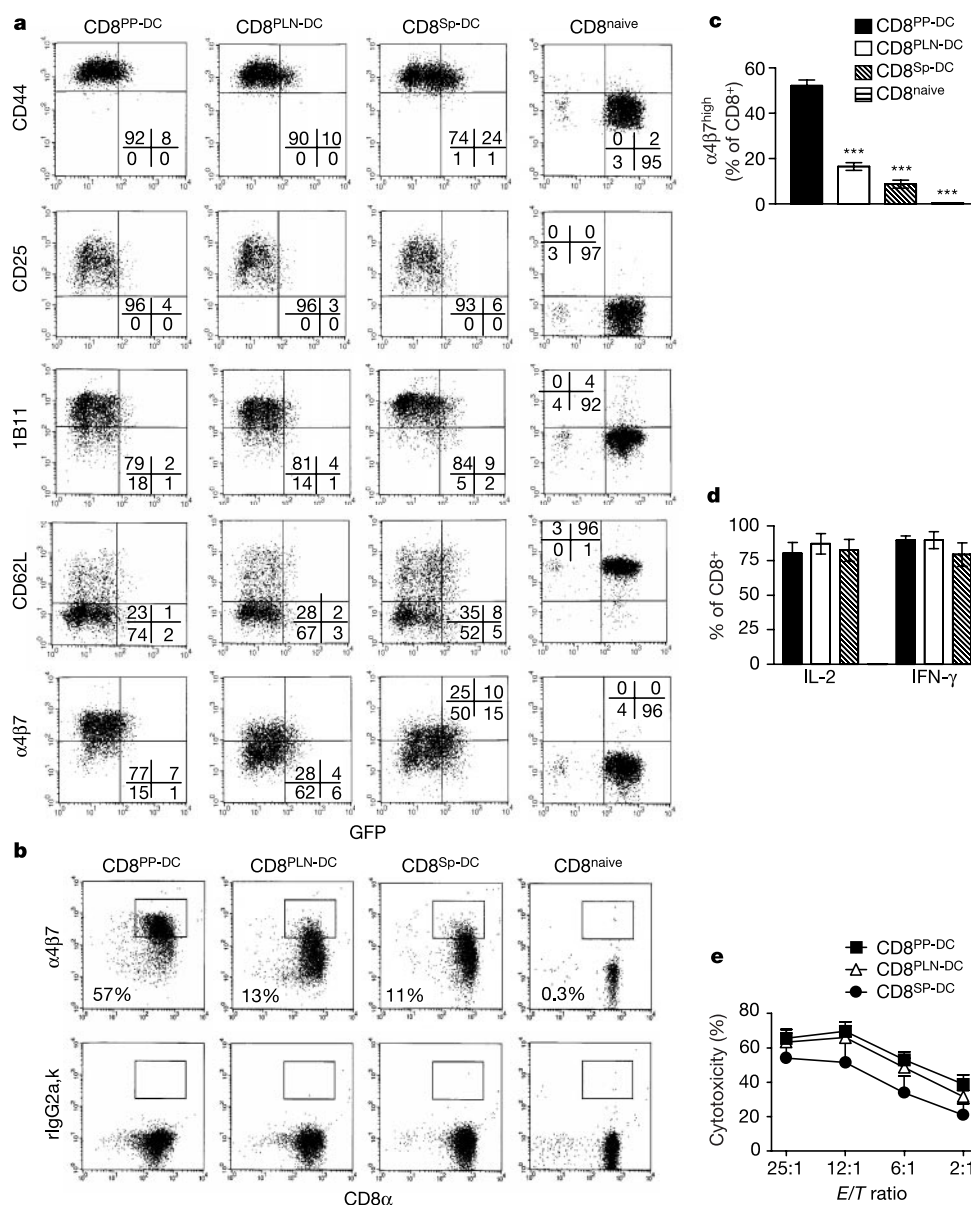


Figure 1 Antigen stimulation of $CD8^+$ cells by DCs from PP, spleen and PLN induces equivalent activation markers and effector activity, but only PP DCs induce high expression of $\alpha 4\beta 7$ integrin. P14 $\times$ T-GFP cells were cultured for 6–7 days with gp33–41-pulsed DCs from different lymphoid organs, and their surface phenotype was compared with naive $CD8^+$ P14 $\times$ T-GFP cells by flow cytometry. **a**, Representative histograms for CD25, CD44, 1B1, L-selectin (CD62L) and $\alpha 4\beta 7$ compared with GFP. **b**, Representative staining of $\alpha 4\beta 7$ heterodimer (top row) and isotype-matched control MAb (bottom row) versus CD8 α . Numbers in histograms reflect the frequency of cells in

the $CD8^+ \alpha 4\beta 7^{high}$ gate (rectangles). **c**, Statistical analysis shows that the frequency of $CD8^+ \alpha 4\beta 7^{high}$ cells was significantly higher in cells stimulated by PP DCs ($n = 22$ for $CD8^{PP-DC}$ and $CD8^{PLN-DC}$, 8 for $CD8^{Sp-DC}$ and 9 for naive; *** $P < 0.001$ compared with $CD8^{PP-DC}$). By contrast, all DC populations induced equivalent effector functions as evidenced by **d**, the production of cytokines induced by phorbol 12-myristate 13-acetate/ionomycin ($n = 3$) and **e**, antigen-specific cytotoxicity ($n = 3$). E/T ratio, effector-to-target ratio.

Table 2 Phenotype of DCs from PP, PLN and spleen

	PP DCs		PLN DCs		Spleen DCs	
	CD8 α^+	CD8 α^-	CD8 α^+	CD8 α^-	CD8 α^+	CD8 α^-
% of DC subset	38 $\pm$ 14	62 $\pm$ 14	61 $\pm$ 14**	39 $\pm$ 14**	48 $\pm$ 8	52 $\pm$ 8
H-2K ^b (MHC class II)	148 $\pm$ 72	122 $\pm$ 61	183 $\pm$ 42	165 $\pm$ 50	142 $\pm$ 22	109 $\pm$ 21
I-A/I-E (MHC class II)	770 $\pm$ 326	808 $\pm$ 303	844 $\pm$ 155	1,256 $\pm$ 310	740 $\pm$ 266	569 $\pm$ 236
CD40	5 $\pm$ 2	3 $\pm$ 2	3 $\pm$ 5	3 $\pm$ 5	4 $\pm$ 2	1 $\pm$ 3
CD80	41 $\pm$ 17	25 $\pm$ 6	22 $\pm$ 5*	21 $\pm$ 7	24 $\pm$ 5	15 $\pm$ 4*
CD86	13 $\pm$ 9	10 $\pm$ 5	12 $\pm$ 5	12 $\pm$ 5	11 $\pm$ 4	6 $\pm$ 2
CD1d (%)	30 $\pm$ 5 (68 $\pm$ 10)	25 $\pm$ 1 (46 $\pm$ 9)	41 $\pm$ 11 (77 $\pm$ 1)	31 $\pm$ 1 (50 $\pm$ 4)	81 $\pm$ 32 (91 $\pm$ 3)	33 $\pm$ 8 (56 $\pm$ 1)
CD48	840 $\pm$ 28	701 $\pm$ 14	1,025 $\pm$ 12	955 $\pm$ 42	784 $\pm$ 109	656 $\pm$ 60
ICAM-1	504 $\pm$ 9	390 $\pm$ 23	500 $\pm$ 35	416 $\pm$ 52	515 $\pm$ 119	260 $\pm$ 39

Data are mean $\pm$ s.d. of MFI (or percentage of marker-positive cells) after gating on CD11c⁺CD8 α^+ or CD11c⁺CD8 α^- cells. $n = 3$ independent DC preparations. * $P < 0.05$, ** $P < 0.01$ compared with corresponding PP DC subset. ICAM-1, intercellular adhesion molecule 1.

As $\alpha 4\beta 7$ and CCR9 are essential intestinal homing receptors^{9,18,22}, we postulated that adoptively transferred CD8^{PP-DC} should home to the gut. To test this, we performed competitive homing experiments by injecting into non-transgenic recipients a mixture of naive T-GFP cells and CD8^{PP-DC} or CD8^{PLN-DC} stained with tetramethylrhodamine isothiocyanate (TRITC), a red fluorophore. The homing index in recipient tissues (the ratio of TRITC⁺ to GFP⁺ cells, corrected for input) was determined 4 h later (Fig. 3a). Compared with naive T cells, CD8^{PP-DC} and CD8^{PLN-DC} homed poorly to lymph nodes, whereas all three T-cell populations had equivalent spleen-homing abilities. The two CTL populations accumulated in the liver and lungs in much larger numbers than naive T cells. This migratory behaviour is typical for naive and effector CD8⁺ cells²³. Interestingly, although CD8^{PP-DC} migrated less than naive T cells to PP (about threefold when corrected for circulating cell counts), CD8^{PP-DC} homed to PP significantly better than CD8^{PLN-DC}. Conversely, CD8^{PLN-DC} retained a higher capacity than CD8^{PP-DC} to home to skin-draining PLN.

We also counted homed T cells in the lamina propria (LP) and among intra-epithelial lymphocytes (IEL) in the small intestine. In most cases, GFP⁺ naive T cells were undetectable in these tissues, whereas CD8^{PP-DC} cells were routinely recovered. Because the absence of naive T cells prohibited us from calculating intestinal homing indices, we carried out further competitive homing experiments with differentially labelled CD8^{PLN-DC} and CD8^{PP-DC} (Fig. 3b). The ratio of CD8^{PP-DC} to CD8^{PLN-DC} was 17 ± 6 times (mean $\pm$ s.e.m.) higher in the small intestine than in the blood. Comparably pronounced differences were also observed when the fluorescent dyes were switched (not shown), or when CD8^{PP-DC} and CD8^{PLN-DC} were generated from OT-I $\times$ RAG^{-/-} mice (Supplementary Fig. S-5c). There was a significant positive correlation between the ratio of $\alpha 4\beta 7$ staining on CD8^{PP-DC} versus CD8^{PLN-DC}, and the homing index in the small intestine (Fig. 3c), indicating that the induction of $\alpha 4\beta 7^{\text{high}}$ T cells by PP DCs is a prerequisite mechanism for effector-cell targeting to that organ. Furthermore, in two independent experiments, a neutralizing MAb to TECK reduced the homing index by 51% compared with an isotype-matched control MAb. Thus, consistent with a recent report⁹, responsiveness to TECK is required for efficient effector-cell homing to the small intestine (Supplementary Fig. S-6).

CD8^{PP-DC} homed significantly more to PP and MLN than did CD8^{PLN-DC} (Fig. 3b), but these differences were comparatively modest. On the other hand, the superior homing ability of CD8^{PP-DC} over CD8^{PLN-DC} was specific for the digestive tract, as their concentration was similar in blood, lung and spleen. Interestingly, despite their pronounced tropism for the small intestine, CD8^{PP-DC} migrated poorly to the colon, similar to CD8^{PLN-DC} (homing index, 1.3 ± 0.2). This observation lends experimental support to the notion that the small and large bowel have distinct homing requirements²⁴.

Previous work has shown that individual DC subsets can bias T-cell differentiation, and DCs from different lymphoid organs can elicit distinct T-cell responses^{25,26}. For example, PLN DCs and PP

DCs polarize CD4⁺ cells preferentially towards a T_H1 and T_H2 (T helper cell) phenotype, respectively^{25,26}. Thus, it was important to assess whether PP DCs exerted a similar effect on CD8⁺ cells, as *in vitro* polarized T_C1 and T_C2 (T cytotoxic) cells have distinct migratory properties²⁷. However, CD8^{PP-DC} and CD8^{PLN-DC} underwent comparable effector differentiation (Fig. 1d, e). Thus, the selective ability of PP DCs to confer gut tropism was not mirrored in a distinct capacity to induce polarized CTL responses.

We also examined the composition of DCs, as normal lymphoid

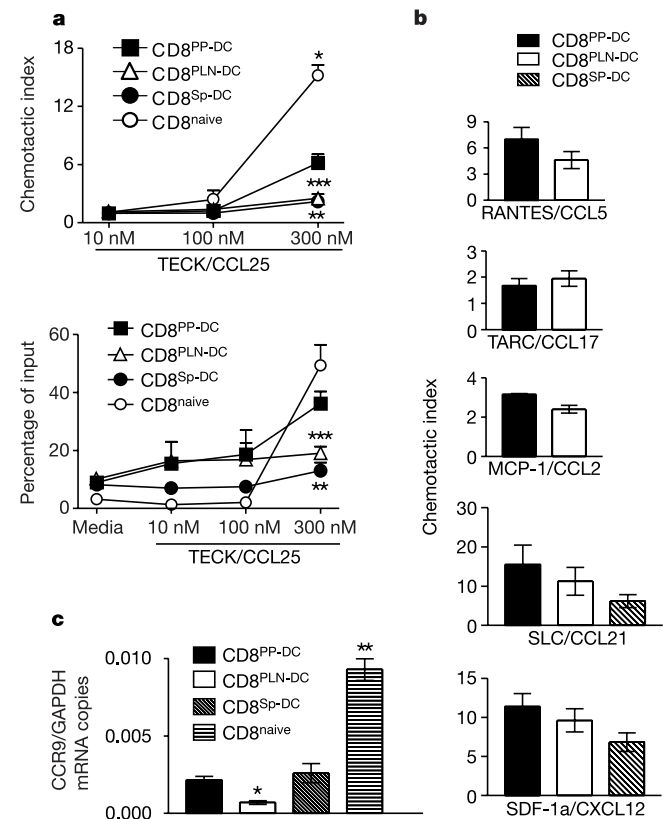


Figure 2 CD8^{PP-DC} are significantly more responsive than CD8^{PLN-DC} or CD8^{SP-DC} to the gut-associated chemokine TECK/CCL25. **a**, P14 $\times$ T-GFP cells were cultured for 6–7 days with gp33–41-pulsed DCs isolated from different lymphoid tissues and compared with naive P14 $\times$ T-GFP cells in Transwell chemotaxis assays using different concentrations of TECK ($n = 6$ –22). Data are shown as chemotactic index; that is, the number of cells migrated towards TECK normalized to the number migrated towards medium (top), and as percentage of input cells recovered from the bottom chamber (bottom). **b**, Migration towards an optimum concentration of RANTES/CCL5, SDF/CXCL12, SLC/CCL21, MCP-1/CCL2 and TARC/CCL17. **c**, Real-time RT-PCR of CCR9 mRNA relative to GAPDH mRNA in CD8⁺ effector cells and purified naive P14 $\times$ T-GFP T cells ($n = 2$ –3). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ compared with CD8^{PP-DC}.

tissues contain several DC subsets with distinct expression patterns of CD8 α and CD11b²⁵. In Flt3-L-treated donors, PP contained fewer CD8 α ⁺ DCs than did PLN (Table 2). However, CD8 α ⁺ DCs were equally represented in PP and spleen, and depletion of CD8 α ⁺ cells from PP DCs and PLN DCs did not alter their differential effect on α 4 β 7 expression and TECK responsiveness by T cells (not shown). All DC populations showed similar levels of co-stimulatory and classical MHC molecules as well as non-classical MHC molecules such as Qa-1 and Qa-2, which have been implicated in intestinal targeting of CD8⁺ recent thymic emigrants²⁸ (not shown). Thus, the mechanism(s) by which PP DCs confer tissue specificity remain(s) to be identified.

After a T cell has been activated in PP, it returns by way of efferent lymph, MLN and the thoracic duct to the blood stream, and then homes to the gut mucosa to establish a first line of defence against recurrent intestinal pathogens⁵. Of note, CD8^{PP-DC} accumulated in non-inflamed gut. Thus, they followed constitutively expressed traffic signals provided, at least in part, by MADCAM-1/ α 4 β 7 and TECK/CCR9. Inflammatory mediators induce a plethora of further recruitment pathways^{1,2,4}. Our analysis shows that DCs from any lymphoid organ prompt CTLs to upregulate an arsenal of traffic molecules to respond to such inflammatory signals. Thus, intestinal inflammation would probably not only amplify CTL recruitment, but may also lessen the requirement for migratory specialization.

Although antigen-induced default changes in common inflammatory traffic molecules on T cells are well established^{1,2,4}, our

findings reveal a novel, tissue-specific aspect of DC-induced effector/memory cell differentiation. We show that PP DCs can target activated CD8⁺ cells to the small intestine, whereas other DCs do not have this ability. We have also observed that PP DCs upregulate α 4 β 7 expression on CD4⁺ cells (J.R.M., M.R.B. and M.R., unpublished observation). As their intestine-specific character was maintained after PP DCs were transferred into tissue culture, it seems unlikely that additional microenvironmental signals (such as stroma cells) are crucial for tissue-specific imprinting, at least in PP. However, local environmental cues probably provide necessary differentiation signals to DCs or DC precursors that enter PP. Whether DCs in other lymphoid organs have effector-targeting properties for associated peripheral tissues remains to be tested.

It will also be interesting to determine whether PP DCs confer gut tropism by selection or instruction. In a scenario governed by selection, T cells initiate random differentiation programmes that result in diverse tissue specificity, and appropriate subsets are then selected by DCs to proliferate and/or survive¹⁰. Alternatively, PP DCs may provide differentiation signals a priori to induce α 4 β 7 and possibly CCR9. In our experiments, DCs rapidly disappeared from co-cultures and were essentially absent after day 3, whereas α 4 β 7^{high} CD8^{PP-DC} arose only after day 4. This observation is more easily reconciled with an instructive mechanism and supports the view that tissue-specific T-cell memory reflects differentiation concurrent with the activation process rather than post-activation selection. □

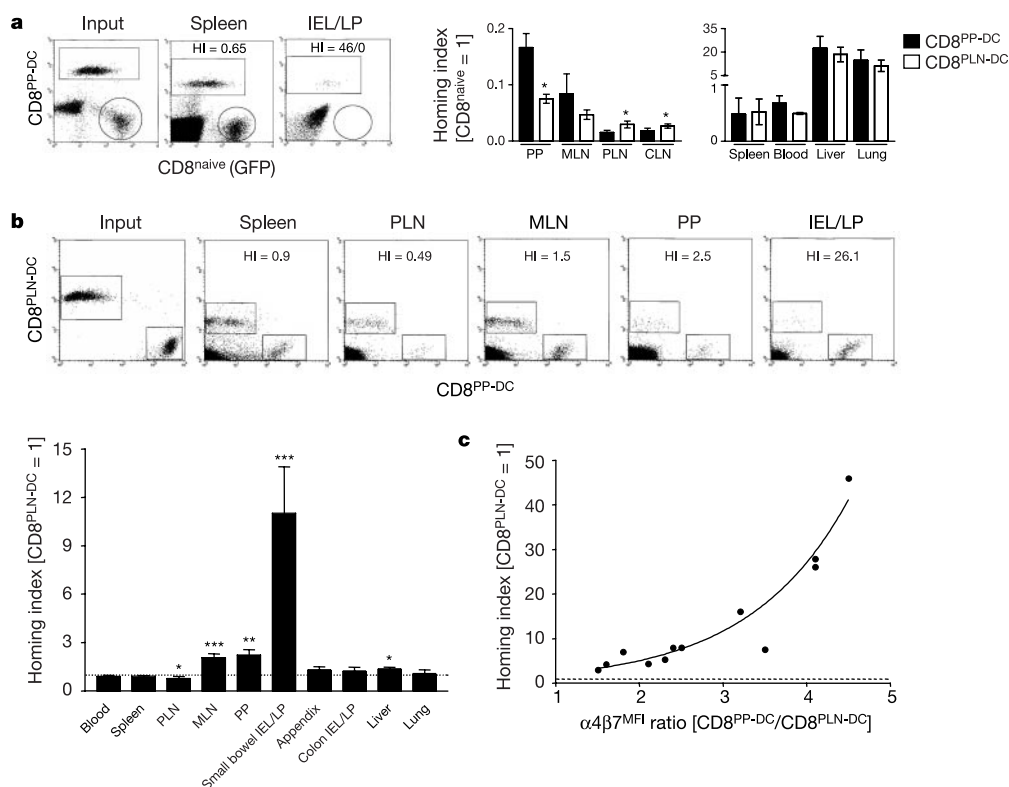


Figure 3 CD8^{PP-DC} home to the small intestine. P14 $\times$ T-GFP cells were cultured for 6–7 days with gp33–41-pulsed DCs isolated from PLN or PP of the same donor pool. **a**, In one series, stimulated cells were labelled with TRITC and mixed with an equal number of T-GFP cells (containing ~35% GFP⁺ naive T cells), and the homing index (HI, the ratio of [TRITC⁺ cells] to [GFP⁺ cells] in recipient blood or tissues divided by the input ratio) was determined in recipients' blood and tissues 4 h after adoptive transfer. FACS histograms (left) show a representative experiment comparing CD8^{PP-DC} and naive T-GFP cells (non-fluorescent input cells are non-T cells among donor splenocytes). Bar graphs (right) show mean $\pm$ s.e.m. of three experiments (*, $P < 0.05$ compared with CD8^{PP-DC}).

b, In a second series, CD8^{PP-DC} and CD8^{PLN-DC} were mixed after staining with calcein AM (or CFSE) and TRITC, respectively, and the HI (the ratio of [calcein⁺ (or CFSE⁺)] to [TRITC⁺], corrected for input) was determined 24 h later. Histograms from a representative experiment and mean $\pm$ s.e.m. of all experiments are shown ($n = 3$ –12; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ compared with HI = 1). **c**, Correlation between the ratios of α 4 β 7 expression on CD8^{PP-DC} compared with CD8^{PLN-DC} and the HI of both populations in small intestine ($r^2 = 0.95$, $P < 0.0001$). The broken line represents HI = 1.

Methods

For a detailed description of methods used in this study, see Supplementary Information.

Mice and reagents

C57BL/6 mice were obtained from Jackson Labs. T-GFP and P14 × T-GFP mice have been described previously⁴. OT-I × RAG^{-/-} mice, ovalbumin-derived SIINFEKL peptide and Flt3-L transfectants were provided by H. Ploegh. Mice were housed in a SPF/VAF (specific-pathogen-free and viral-antibody-free) facility and used between 8–16 weeks of age in accordance with National Institutes of Health guidelines. MABs and cytokine staining kits were from Pharmingen. LCMV gp_{33–41} peptide (KAVYNFATC) was from Biosource. Murine rIL-2 and chemokines were from R&D Systems.

DC isolation

Female C57BL/6 mice were injected subcutaneously with B16 melanoma cells secreting Flt3-L¹¹. After 12–17 days, mice were killed and single-cell suspensions were generated by digesting PB, PLN and spleen with collagenase D and DNase I (Roche). DCs were immunomagnetically isolated by negative selection using MABs to B220, CD3, CD19, Pan-NK, Ter-119, Ly-6G and Thy-1²⁹, and goat anti-rat IgG microbeads (Miltenyi Biotec). Negatively selected DCs were pulsed with 10 μg ml⁻¹ gp_{33–41} or SIINFEKL peptide in IMDM with standard supplements and 10% FBS (complete IMDM), and used immediately in co-cultures. For some experiments, negatively selected DCs were stained with fluorescein isothiocyanate (FITC)-anti-CD11c and a cocktail of phycoerythrin (PE)-labelled lineage-specific antibodies, and sorted into CD11c⁺ and CD11c⁻ fractions using a FACS Vantage cell sorter (Becton Dickinson). Sorted cells were pulsed with peptide and used as described above.

T-cell isolation and T-cell/DC co-cultures

Naive CD8⁺ P14 × T-GFP and OT-I × Rag^{-/-} T cells were purified from splenocytes by red blood cell lysis ('ACK' lysis buffer) followed by negative immunomagnetic selection using MABs to B220, I-A/I-E, CD4, CD19, Pan-NK, Ter-119 and Ly-6G, as described above. T cells (90–95% CD8⁺, L-selectin^{high}, CD4^{low}, GFP^{high}) were co-cultured with 1 × 10⁶ peptide-pulsed DCs (ratio of T cells to DCs, 1:1 or 2:1) from different lymphoid organs in 3 ml of complete IMDM using 12-well tissue-culture treated plates (Falcon). On day 3, 2 ml of complete IMDM and a single dose of 10 ng ml⁻¹ IL-2 were added to maintain viability, and 2 ml of complete IMDM were replaced daily thereafter. Cells were used after 6–7 days in culture. For some experiments, naive P14 × T-GFP cells were stimulated with and without the addition of sorted CD11c⁺ or CD11c⁻ fractions. For this, 12-well plates were coated with anti-CD3 and anti-CD28, and CD8⁺ P14 × T-GFP cells were added alone or together with sorted CD11c⁺ or CD11c⁻ cells (ratio of T cells to sorted cells, 1:1 or 2:1). After 3 days in culture, T cells were transferred to plates without antibodies and maintained as described above.

Characterization of CTL phenotype

The surface phenotype and intracellular cytokine production of differentiated T-cell populations were analysed on a FACSscan flow cytometer (Becton Dickinson) using standard techniques. To assess antigen-specific cytotoxicity, a 6 h chromium release assay was carried out as described³⁰. To measure CCR9 mRNA levels, the ratio of message levels for CCR9 and the housekeeping gene GAPDH (glyceraldehyde-3-phosphate dehydrogenase) was determined by real-time RT-PCR (polymerase chain reaction with reverse transcription) using an iCycler (Bio-Rad) and the SYBR Green DNA PCR Core Kit (Applied Biosystems).

Chemotaxis assay

A standard 90 min chemotaxis assay with Transwell filters (pore size, 5 μm) was used to assess CTL responsiveness to different chemokines²³. Chemotactic indices were calculated as the number of migrated cells in wells containing chemokine, divided by the number of migrated cells in wells containing medium alone. The percentage of migrated cells relative to cells in the input was also determined.

Homing assay

Competitive homing experiments of TRITC-labelled CD8^{PLN-DC} and CD8^{PP-DC} cells and naive T-GFP splenocytes were performed as described elsewhere²³. Briefly, 2 × 10⁷ CTLs were mixed with the same number of T-GFP cells and injected intravenously into non-transgenic recipients. An aliquot was saved to assess the input ratio (IR = [TRITC⁺]_{input}/[GFP⁺]_{input}). After 4 h, recipient tissues were harvested to measure TRITC⁺/GFP⁺ ratios by flow cytometry. The homing index, HI, was calculated as the ratio of [TRITC⁺]_{tissue}/[GFP⁺]_{tissue} to IR. To compare the migration of CD8^{PLN-DC} and CD8^{PP-DC}, one population was labelled with TRITC (Molecular Probes) and the other with calcein AM or CFSE (Molecular Probes). 2 × 10⁷ cells from each population were mixed and used for competitive homing experiments. In two experiments, mice were injected with 100 μg rIgG2b or anti-TECK/CCl25 (R&D Systems, clone 89818) 4 h before and during CTL injection. Recipients were sacrificed after 18 h and lymphocytes from LP and IEL were obtained as described³. The HI was calculated as the ratio of [CD8^{PP-DC}]_{tissue}/[CD8^{PLN-DC}]_{tissue} to [CD8^{PP-DC}]_{input}/[CD8^{PLN-DC}]_{input}.

Statistical analysis

Data are presented as mean ± s.e.m. and were analysed using either the Kruskal-Wallis test with Dunn's post-test or one-way ANOVA with Bonferroni correction, as appropriate. Homing indices were tested versus HI = 1 using a one-sample Wilcoxon-signed rank test. Significance was set at *P* < 0.05.

Received 14 January; accepted 12 March 2003; doi:10.1038/nature01726.

- Butcher, E. C., Williams, M., Youngman, K., Rott, L. & Briskin, M. Lymphocyte trafficking and regional immunity. *Adv. Immunol.* **72**, 209–253 (1999).
- von Andrian, U. H. & Mackay, C. R. T-cell function and migration. Two sides of the same coin. *N. Engl. J. Med.* **343**, 1020–1034 (2000).
- Masopust, D., Vezys, V., Marzo, A. L. & Lefrançois, L. Preferential localization of effector memory cells in nonlymphoid tissue. *Science* **291**, 2413–2417 (2001).
- Weninger, W., Manjunath, N. & von Andrian, U. H. Migration and differentiation of CD8⁺ T cells. *Immunol. Rev.* **186**, 221–233 (2002).
- Guy-Grand, D., Griscelli, C. & Vassalli, P. The mouse gut T lymphocyte, a novel type of T cell. Nature, origin, and traffic in mice in normal and graft-versus-host conditions. *J. Exp. Med.* **148**, 1661–1677 (1978).
- Kantele, A., Zivny, J., Hakkinen, M., Elson, C. O. & Mestecky, J. Differential homing commitments of antigen-specific T cells after oral or parenteral immunization in humans. *J. Immunol.* **162**, 5173–5177 (1999).
- Campbell, D. J. & Butcher, E. C. Rapid acquisition of tissue-specific homing phenotypes by CD4⁺ T cells activated in cutaneous or mucosal lymphoid tissues. *J. Exp. Med.* **195**, 135–141 (2002).
- Hamann, A., Andrew, D. P., Jablonski-Westrich, D., Holzmann, B. & Butcher, E. C. Role of α₄-integrins in lymphocyte homing to mucosal tissues *in vivo*. *J. Immunol.* **152**, 3282–3293 (1994).
- Svensson, M. *et al.* CCL25 mediates the localization of recently activated CD8αβ⁺ lymphocytes to the small-intestinal mucosa. *J. Clin. Invest.* **110**, 1113–1121 (2002).
- Davenport, M. P., Grimm, M. C. & Lloyd, A. R. A homing selection hypothesis for T-cell trafficking. *Immunol. Today* **21**, 315–317 (2000).
- Shi, G. P. *et al.* Cathepsin S required for normal MHC class II peptide loading and germinal center development. *Immunity* **10**, 197–206 (1999).
- Nakache, M., Berg, E. L., Streeter, P. R. & Butcher, E. C. The mucosal vascular addressin is a tissue-specific endothelial cell adhesion molecule for circulating lymphocytes. *Nature* **337**, 179–181 (1989).
- Berlin, C. *et al.* α₄ integrins mediate lymphocyte attachment and rolling under physiologic flow. *Cell* **80**, 413–422 (1995).
- Schott, E. & Ploegh, H. L. Mouse MHC class I tetramers that are unable to bind to CD8 reveal the need for CD8 engagement in order to activate naive CD8 T cells. *Eur. J. Immunol.* **32**, 3425–3434 (2002).
- Stagg, A. J., Kamm, M. A. & Knight, S. C. Intestinal dendritic cells increase T cell expression of α₄β₇ integrin. *Eur. J. Immunol.* **32**, 1445–1454 (2002).
- Andrew, D. P. *et al.* Distinct but overlapping epitopes are involved in α₄β₇-mediated adhesion to vascular cell adhesion molecule-1, mucosal addressin-1, fibronectin, and lymphocyte aggregation. *J. Immunol.* **153**, 3847–3861 (1994).
- Hynes, R. O. Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673–687 (2002).
- Zabel, B. A. *et al.* Human G protein-coupled receptor GPR-9-6/CC chemokine receptor 9 is selectively expressed on intestinal homing T lymphocytes, mucosal lymphocytes, and thymocytes and is required for thymus-expressed chemokine-mediated chemotaxis. *J. Exp. Med.* **190**, 1241–1256 (1999).
- Zaballos, A., Gutierrez, J., Varona, R., Ardavin, C. & Marquez, G. Cutting edge: identification of the orphan chemokine receptor GPR-9-6 as CCR9, the receptor for the chemokine TECK. *J. Immunol.* **162**, 5671–5675 (1999).
- Carramolino, L. *et al.* Expression of CCR9 β-chemokine receptor is modulated in thymocyte differentiation and is selectively maintained in CD8⁺ T cells from secondary lymphoid organs. *Blood* **97**, 850–857 (2001).
- Iwasaki, A. & Kelsall, B. L. Localization of distinct Peyer's patch dendritic cell subsets and their recruitment by chemokines macrophage inflammatory protein (MIP)-3α, MIP-3β, and secondary lymphoid organ chemokine. *J. Exp. Med.* **191**, 1381–1394 (2000).
- Wagner, N. *et al.* Critical role for β₇ integrins in formation of the gut-associated lymphoid tissue. *Nature* **382**, 366–370 (1996).
- Weninger, W., Crowley, M. A., Manjunath, N. & von Andrian, U. H. Migratory properties of naive, effector, and memory CD8⁺ T cells. *J. Exp. Med.* **194**, 953–966 (2001).
- Kunkel, E. J. & Butcher, E. C. Chemokines and the tissue-specific migration of lymphocytes. *Immunity* **16**, 1–4 (2002).
- Kelsall, B. L., Biron, C. A., Sharma, O. & Kaye, P. M. Dendritic cells at the host–pathogen interface. *Nature Immunol.* **3**, 699–702 (2002).
- Banchereau, J. *et al.* Immunobiology of dendritic cells. *Annu. Rev. Immunol.* **18**, 767–811 (2000).
- Xie, H., Lim, Y. C., Lusinskas, F. W. & Lichtman, A. H. Acquisition of selectin binding and peripheral homing properties by CD4⁺ and CD8⁺ T cells. *J. Exp. Med.* **189**, 1765–1776 (1999).
- Maurice, M. M., Gould, D. S., Carroll, J., Vugmeyster, Y. & Ploegh, H. L. Positive selection of an MHC class-I restricted TCR in the absence of classical MHC class I molecules. *Proc. Natl Acad. Sci. USA* **98**, 7437–7442 (2001).
- Suss, G. & Shortman, K. A subclass of dendritic cells kills CD4 T cells via Fas/Fas-ligand-induced apoptosis. *J. Exp. Med.* **183**, 1789–1796 (1996).
- Manjunath, N. *et al.* A transgenic mouse model to analyse CD8⁺ effector T cell differentiation *in vivo*. *Proc. Natl Acad. Sci. USA* **96**, 13932–13937 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to H. Ploegh for providing reagents and mice. We thank I. Dodge, B. Harnisch, G. Cheng, N. Barteneva, C. Halin, E. Schott and S. Feske for advice and technical assistance. J.R.M. is indebted to Ingrid Ramos for constant support. This work was supported by grants from NIH to U.H.v.A.; from Fondecyt to J.R.M., M.R.B. and M.R.; and from MIFAB (financed in part by Ministerio de Planificación y Cooperación, Chile) to M.R. W.W. is a fellow of the Max Kade Foundation. J.R.M. was supported in part by fellowships from Fundación Andes and the Pew Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to U.H.v.A. (uva@cbr.med.harvard.edu).

The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA

Hui Zhang, Bin Yang, Roger J. Pomerantz, Chune Zhang, Shyamala C. Arunachalam & Ling Gao

The Dorrance H. Hamilton Laboratories, Center for Human Virology and Biodefense, Division of Infectious Diseases and Environmental Medicine, Department of Medicine, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA

High mutation frequency during reverse transcription has a principal role in the genetic variation of primate lentiviral populations. It is the main driving force for the generation of drug resistance and the escape from immune surveillance. G to A hypermutation is one of the characteristics of primate lentiviruses, as well as other retroviruses, during replication *in vivo* and in cell culture^{1–6}. The molecular mechanisms of this process, however, remain to be clarified. Here, we demonstrate that CEM15 (also known as apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3G; APOBEC3G)^{7,8}, an endogenous inhibitor of human immunodeficiency virus type 1 (HIV-1) replication, is a cytidine deaminase and is able to induce G to A hypermutation in newly synthesized viral DNA. This effect can be counteracted by the HIV-1 virion infectivity factor (Vif). It seems that this viral DNA mutator is a viral defence mechanism in host cells that may induce either lethal hypermutation or instability of the incoming nascent viral reverse transcripts, which could account for the Vif-defective phenotype. Importantly, the accumulation of CEM15-mediated non-lethal hypermutation in the replicating viral genome could potentially contribute to the genetic variation of primate lentiviral populations.

HIV-1 Vif protein is required for viral replication *in vivo* and in some 'non-permissive' cells, such as peripheral blood mononuclear cells, macrophages and H9 T cells^{9–11}. The *vif*-defective viruses (Δ vif) from non-permissive cells cannot complete reverse transcription, or the newly synthesized DNA cannot exist in the target cells for a significant time period^{12–14}. Recently, it has been demonstrated that CEM15 is an endogenous inhibitor of HIV-1 that exists only in non-permissive cells. Its inhibitory effect on HIV-1 replication can be counteracted by HIV-1 Vif protein⁷. As Vif binds to HIV-1 RNA in the cytoplasm of virus-producing cells^{15–17}, we investigated whether CEM15, which shares significant homology with some other cytidine deaminases that edit RNA, could also edit HIV-1 genomic or spliced RNA. We have sequenced the nearly full-length genomic RNA (>98%) of HIV-1 from the Δ vif virions generated from H9 T cells by polymerase chain reaction with reverse transcription (RT-PCR) techniques (primer pairs are listed in Supplementary Table S1). Compared with the sequence of pNL4-3 Δ vif DNA, the change of genomic RNA in the virions is not significant. We have found an A to G change at positions 2257 and 3608, and a G to T change at position 9418 (data not shown). We have also sequenced several spliced HIV-1(NL4-3 Δ vif) RNA in H9 cells (Supplementary Fig. S1 and Table S1); however, no mutations were demonstrated.

CEM15, as well as some other homologues such as AID and APOBEC1, can function as a DNA mutator in *Escherichia coli*^{18,19}. As both CEM15 and Vif can be packaged into HIV-1 virions^{7,17}, we then investigated whether the newly synthesized HIV-1 DNA is a substrate of CEM15. We first compared the sequence of newly synthesized viral DNA in the presence or absence of Vif. The newly synthesized DNA of Δ vif viruses from H9 cells have G to A

substitutions in all the sequences (U3-R-U5, V3 and PR regions; primer pairs are listed in Supplementary Table S1) that we have analysed (Fig. 1a, row 2 of top panel, and data not shown). The mutation frequency at various regions is similar (data not shown). Notably, the pattern of G to A hypermutation in Δ vif viruses from H9 cells is consistent with the pattern of G to A hypermutation occurring during wild-type virus passage in cell culture and during *in vivo* viral growth^{3,5,6}. They both preferentially take place in a GpA or GpG dinucleotide context or in a group of G nucleotides. Besides the G to A substitution, other substitutions have also been found but with much lower frequency. Of note, as C to T substitution is not significantly higher than the rest of substitutions in the analysed regions, it has been classified into the category of 'other substitutions'. Conversely, the newly synthesized viral DNA from endogenous reverse transcription, which was driven by deoxyribonucleoside triphosphates (dNTPs) at equal amount, also has similar mutations, indicating that all the necessary factors have been packaged within the virions for this process, and that the G to A substitution is not dependent on the imbalance of dNTP pools (Fig. 1a, row 3 of top panel). The mutation frequency in viral DNA isolated from the intravirion reverse transcripts at 4 h after endogenous reverse transcription (ERT) initiation and from intracellular reverse transcripts at 12 h after infection is quite similar (data not shown). Conversely, no G to A substitution was found in Δ vif viruses from permissive SupT1 cells and the wild-type viruses generated from H9 or SupT1 cells (Fig. 1a, top panel, and 1b). Furthermore, we have also examined the effect of Vif on hypermutation in long-term cell culture by passing the wild-type or Δ vif viruses in C8166, a T-cell line that is semi-permissive for Δ vif viruses²⁰ and that expresses small amounts of CEM15 (ref. 7 and data not shown). We found that the G to A hypermutation in the Δ vif viruses is markedly higher than that in the wild-type viruses (Fig. 1a, bottom panel, and 1c).

We then examined the effect of CEM15 on G to A substitution. 293T cells, which are permissive cells and in which no endogenous CEM15 is expressed⁷, were transfected with wild-type or Δ vif HIV-1 DNA in the presence or absence of CEM15. After 48 h, the viruses in the supernatants were then allowed to infect C8166 cells. We found that the newly synthesized DNA of Δ vif viruses with CEM15 contained markedly more G to A hypermutations than that in the newly synthesized DNA from Δ vif viruses without CEM15, or from wild-type viruses (Fig. 2a, top panel, and 2b). Again, the endogenous reverse transcripts of Δ vif viruses with CEM15 also contained G to A hypermutations (Fig. 2a, row 5 of top panel). Similar phenomena were also found in the newly synthesized viral DNA of Δ vif viruses that were generated from the SupT1 T cells containing retrovirally transduced CEM15 (Fig. 2b).

However, it is clear that the G to A hypermutation still occurs in the presence of the *vif* gene when the viruses replicate in the cell culture for several passages^{3,21}. We then examined the effect of CEM15 on the mutation frequency of wild-type viruses in long-term cultures. The CEM15 gene was transduced into SupT1 cells by a Moloney murine leukaemia virus (MMLV) retroviral vector. The wild-type viruses were then passaged in the SupT1 cells, with or without CEM15 expression. After four passages, the viral DNA was amplified and sequenced. We found that wild-type viruses passaged in SupT1 cells containing CEM15 have significantly more G to A hypermutations in their genomes than those in SupT1 cells without CEM15 (Fig. 2a, bottom panel, and 2c). This result further demonstrated that CEM15 is able to induce G to A hypermutation and could be responsible, at least in part, for the hypermutation in the long-term culture of wild-type viruses. Less Vif incorporation into virions in chronic infection might decrease its counteracting effects on CEM15 in the virions (Fig. 2d)²². It is notable that CEM15 does not significantly affect the dNTP pools in SupT1 cells and C8166

cells (Supplementary Fig. S2), excluding the possibility that CEM15 induces the imbalances in the dNTP pools in cells.

To verify that CEM15 has cytidine deaminase activity, glutathione *S*-transferase (GST)–CEM15 fusion protein was purified from *E. coli*. Figure 3 illustrates that CEM15 indeed has cytidine deaminase activity that can be inhibited with tetrahydrouridine (THU). By sequence alignment, it is predicted that CEM15 contains two zinc finger domains⁷. The similar zinc finger domains in other cytidine deaminases have a principal role in cytidine deaminase activity²³. To verify that the zinc finger domains in CEM15 are important for cytidine deaminase, several mutations at zinc finger domains were generated. Figure 3b demonstrates that the mutants at either zinc domains have significantly decreased cytidine deaminase activity. To determine whether purified GST–CEM15 contains the enzymatic activity to edit viral RNA, DNA, or DNA–RNA duplexes, PCR-amplified DNA, *in vitro*-transcribed RNA, or *in*

vitro reverse-transcribed RNA–DNA duplexes were incubated with GST–CEM15, respectively. All the nucleic acids were generated from the same HIV-1 fragment (nucleotides 2134–2594). After incubation, the DNA, RNA or RNA–DNA duplexes were used as the templates for PCR or RT–PCR amplification. The PCR products were then inserted into the pGEM-T vector and sequenced. No mutation was found (data not shown). It is probable that, similar to APOBEC1, CEM15 requires additional factor(s) to perform editing^{23,24}.

The CEM15 mutants were then co-transfected with pNL4-3Δvif into 293T cells to test their functional activities. After 48 h, the viruses in the supernatants were collected. The virions, normalized for quantity, were then used to infect C8166 cells or HLCD4-CAT cells. Twelve hours after infection, the newly synthesized DNA in C8166 cells was extracted, followed by PCR amplification and sequencing. The lysate of HLCD4-CAT cells was collected for CAT

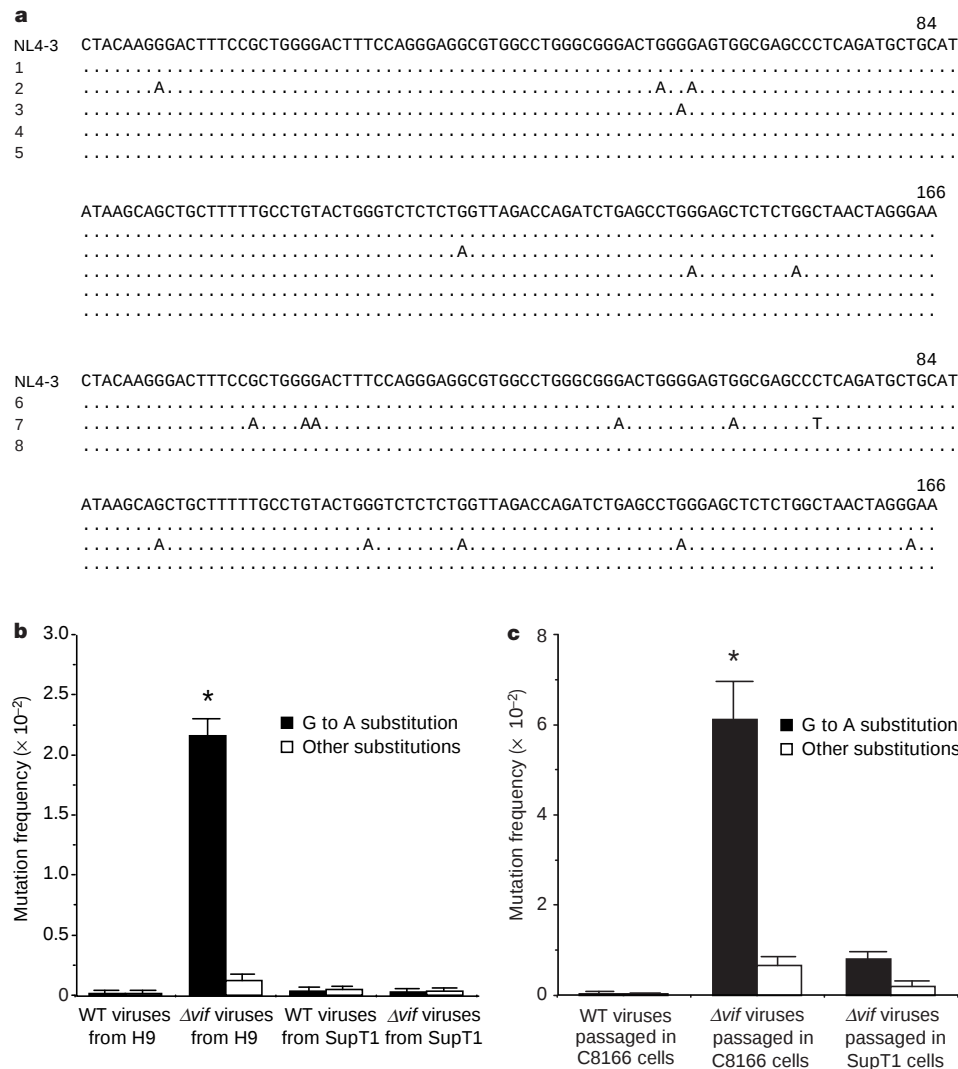


Figure 1 G to A hypermutation in viral DNA of Δ vif viruses from non-permissive cells or semi-permissive cells. **a**, The DNA sequence alignments in the U3-R region. The clones with the most-identical sequences from the total clones ($n = 8$) are listed. The newly synthesized viral DNA are listed in the top panel. The rows are: (1) wild-type viruses from H9 cells to infect C8166 cells; (2) Δ vif viruses from H9 cells to infect C8166 cells; (3) Δ vif viruses from H9 cells to process ERT; (4) wild-type viruses from SupT1 cells to infect C8166 cells; (5) Δ vif viruses from SupT1 cells to infect C8166 cells. After two passages, the viral DNA in C8166 cells was sequenced (bottom panel). The rows are: (6) wild-type

viruses; (7) Δ vif viruses; (8) Δ vif viruses passed in SupT1 cells. **b**, Comparison of mutation frequency of G to A in the newly synthesized DNA of viruses generated from H9 or SupT1 cells (301 nucleotides in U3-R-U5 region, $n = 8$). The G to A mutation frequency of Δ vif viruses from H9 cells to infect C8166 cells is significantly higher than that of others (asterisk, $P < 0.001$, *t*-test). WT, wild type. **c**, Comparison of the G to A mutation frequency in the viral DNA after passage in C8166 cells ($n = 8$). The G to A mutation frequency of Δ vif viruses passed in C8166 is significantly higher than that of wild-type viruses (asterisk, $P < 0.001$, *t*-test).

assay after 48 h. Figure 4 shows that, compared with wild-type CEM15, CEM15 mutants that contain less cytidine deaminase activity could not decrease the infectivity of Δvif viruses in permissive HLC4-CAT cells (Fig. 4a). Furthermore, the frequency of G to A hypermutation in the newly synthesized DNA induced by the mutants is also much lower than that induced by wild-type CEM15 (Fig. 4b). These data suggest that CEM15 induces G to A hypermutation in the newly synthesized viral DNA through its cytidine deaminase activity.

Our data have indicated that the genomic RNA of Δvif viruses from non-permissive cells does not contain the G to A substitution, and CEM15 does not cause a significant imbalance of dNTPs in the

cells. We have also found that, in the presence of CEM15, the product of endogenous reverse transcription of Δvif viruses at high and equal concentration of dNTPs (1 mM) contains the G to A hypermutation. Furthermore, the hypermutation preferentially takes place in a GpA or GpG dinucleotide context or in a group of G nucleotides, which does not occur when the imbalance of the dTTP/dCTP ratio induces dC to dT substitution. On the basis of these observations, we propose that CEM15 may directly deaminate dC in the newly synthesized viral DNA and convert it to dU. The dC to dU conversion in the minus strand DNA of Δvif viruses would further be base-pair-matched by dA during plus-stranded viral DNA synthesis, generating G to A hypermutation in the newly

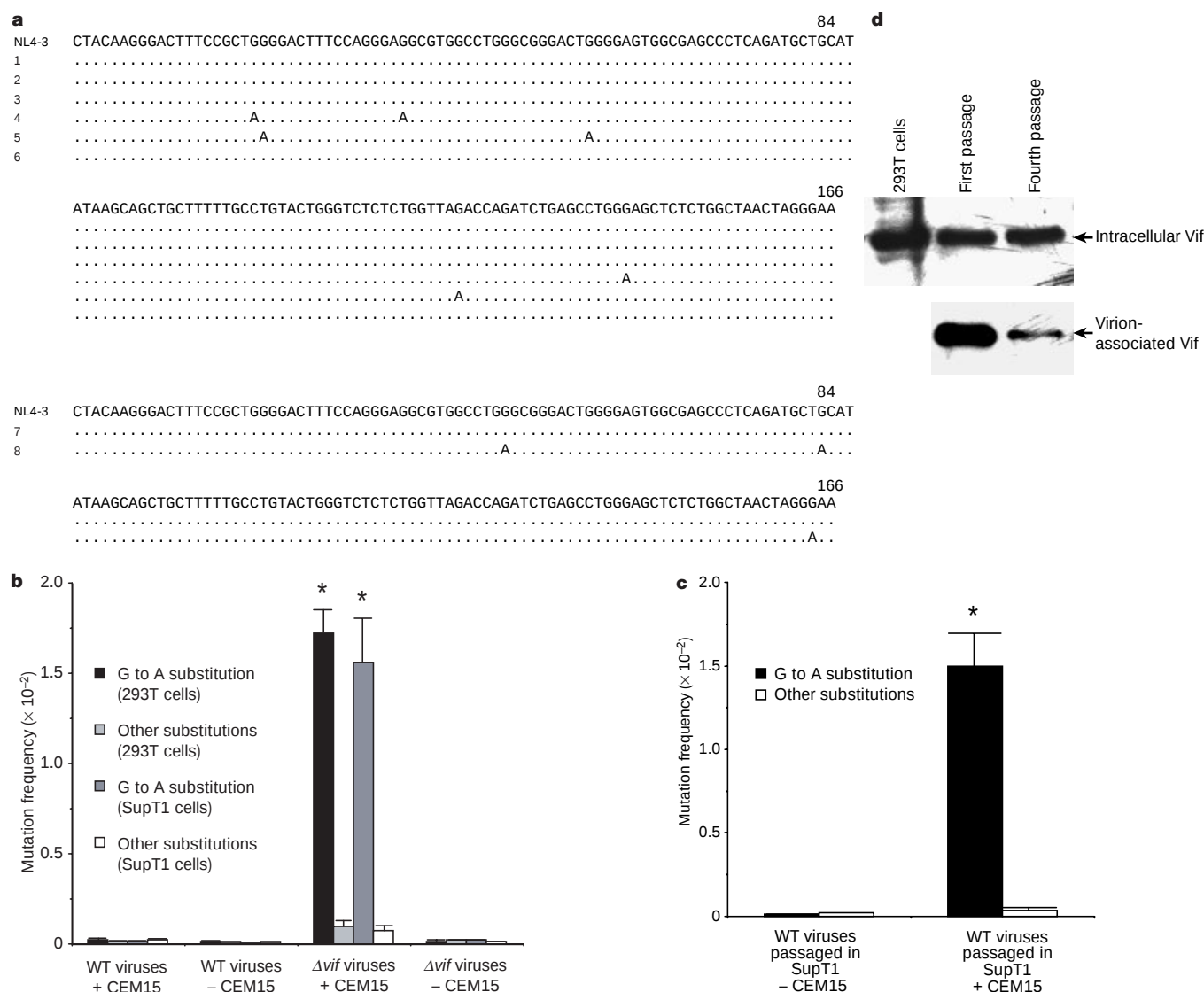


Figure 2 G to A hypermutation in newly synthesized DNA of viruses generated from cells containing CEM15. **a**, DNA sequence alignment in HIV-1 U3-R region. The clones with the most-identical sequences from the total clones ($n = 8$) are listed. Viruses were from 293T (top) cells or passaged in SupT1 (bottom) cells. In the top panel, the rows are: (1) wild-type viruses, with CEM15, to infect C8166 cells; (2) wild-type viruses, with CEM15, to process ERT; (3) wild-type viruses, without CEM15, to infect C8166 cells; (4) Δvif viruses, with CEM15, to infect C8166 cells; (5) Δvif viruses, with CEM15, to process ERT; (6) Δvif viruses, without CEM15, to infect C8166 cells. In the bottom panel, the rows are: (7) wild-type viruses, without CEM15; (8) wild-type viruses, with CEM15. **b**, Comparison of G to A mutation frequency in newly

synthesized DNA (301 nucleotides in U3-R-U5 region, $n = 8$). The G to A mutation frequency of Δvif viruses from 293T or SupT1 cells containing CEM15 is significantly higher than that of others (asterisk, $P < 0.001$, t -test). **c**, Comparison of G to A mutation frequency in the viral DNA after passage ($n = 8$). The G to A mutation frequency of wild-type viruses passaged in SupT1 cells containing CEM15 is significantly higher than that in SupT1 cells without CEM15 (asterisk, $P < 0.001$, t -test). **d**, The wild-type viruses, after various passages in SupT1 cells containing CEM15, were purified²⁹. After normalization with HIV-1 p24, the virion-associated Vif protein was detected by western blot¹⁵.

synthesized viral DNA. If integrated, the Δvif proviral DNA containing the lethal hypermutation might encode 'pre-mature stop' or mutated viral proteins, therefore generating defective viruses. Conversely, the deaminated Δvif viral DNA may become unstable in the cytoplasm. One of the possibilities is that the uracil-DNA glycosylase (UDG) enzyme, which has been packaged into the HIV-1 virions, could excise the dU from the DNA. Accordingly, a break point in the minus strand of viral DNA might occur, which may no longer be able to serve as the template for full-length viral DNA synthesis and thus would gradually be degraded.

Previous work has found that G to A hypermutation cannot be reproduced in a cell- and virus-free reverse transcription reaction when the concentrations and ratios of dNTPs are similar to those in living cells. When the dTTP/dCTP ratio rises to 10,000 to 1, dC to dT substitution occurs²⁵. Furthermore, there is evidence suggesting that G to A hypermutation occurs in thymidine-treated U937-2 cells²⁶. Here, we have not only indicated that G to A hypermutation occurs in the newly synthesized DNA of Δvif viruses generated from the non-permissive cells, but we have also

shown that CEM15 is responsible for this phenomenon. Furthermore, we have demonstrated that CEM15 induces non-lethal hypermutation in the viral DNA of wild-type viruses passaged in a long-term culture. Therefore, we propose that CEM15 could be responsible for the hypermutation observed *in vivo* and in cell culture³⁻⁶.

It has been reported that interferon-inducible, double-stranded RNA-specific adenosine deaminase can also extensively edit the

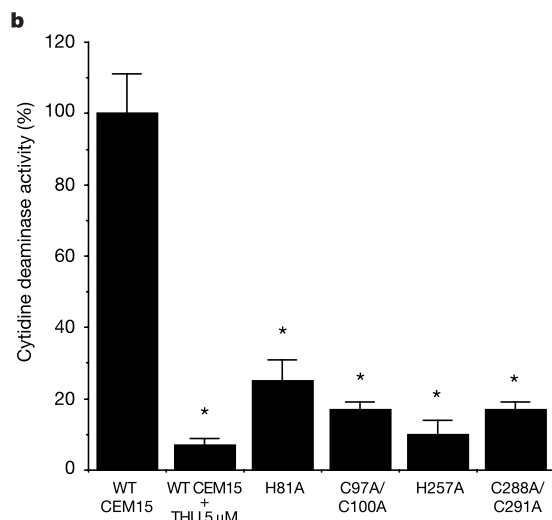
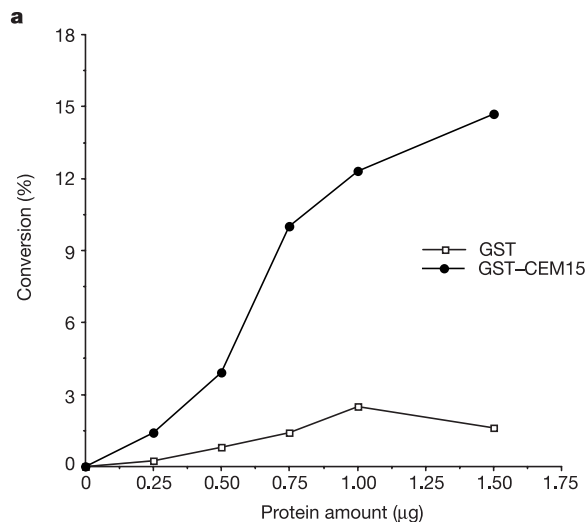


Figure 3 CEM15 has cytidine deaminase activity *in vitro*. **a**, GST-CEM15, but not GST alone, converts deoxycytidine to deoxyuridine *in vitro* in a concentration-dependent manner. This figure represents five independent experiments. **b**, Mutations at zinc finger domains significantly decrease cytidine deaminase activity (asterisk, $P < 0.001$, *t*-test). THU significantly inhibits cytidine deaminase activity (asterisk, $P < 0.001$, *t*-test).

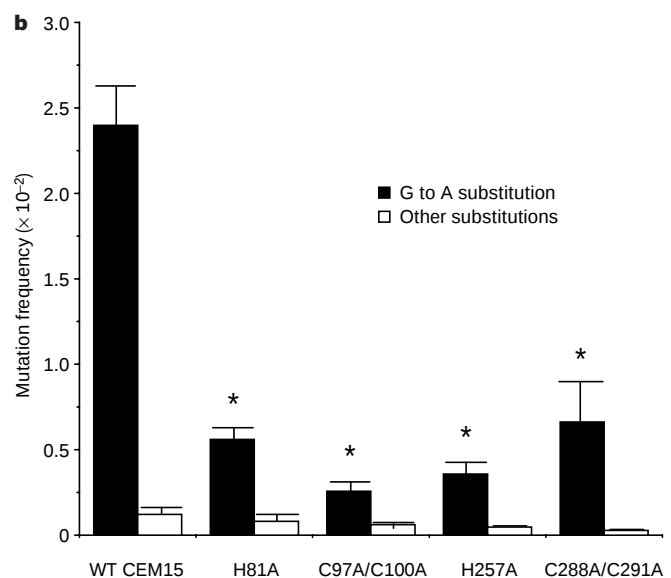
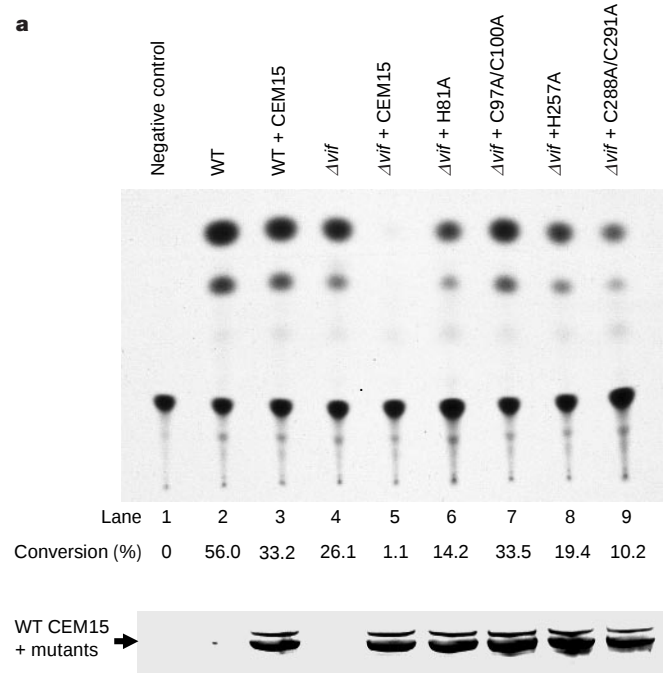


Figure 4 CEM15 without cytidine deaminase activity can neither inhibit the infectivity nor induce hypermutation in the newly synthesized DNA of Δvif viruses. pNL4-3 or pNL4-3 Δvif were transfected into 293T cells with CEM15 or CEM15 mutants. **a**, The viruses were used to infect HLCD4-CAT cells. CAT assays were performed. This figure represents three independent experiments. **b**, The newly synthesized viral DNA in C8166 cells was amplified by PCR and sequenced. The G to A mutation frequency of Δvif viruses from 293T cells containing mutant CEM15 genes is significantly lower than that containing wild-type CEM15 (asterisk, $P < 0.01$, *t*-test).

RNA genome of some RNA viruses in virus-producing cells^{27,28}. We have now demonstrated that G to A hypermutation in newly synthesized reverse transcripts might be induced by a similar host defence factor. Further investigations are needed to determine the precise molecular mechanism(s) of this unique interaction of primate lentiviruses with the host cells. □

Methods

Plasmid constructions, mutagenesis and protein purification

The CEM15 gene with a Flag-tag sequence at its 3' terminus was amplified from the messenger RNA of H9 cells through RT-PCR, and the sequence was confirmed. It was then cloned into various vectors: pcDNA3 for transfection into 293T cells; pGEX for GST fusion protein expression; and pSLX-CMV for retroviral (MLLV) transduction into SupT1 cells. CEM15 mutations were generated by a PCR-based mutagenesis approach¹⁵. The GST, GST-CEM15 and other GST fusion CEM15 mutant proteins were produced according to previously described methods¹⁵.

Transfections

The 293T or T-lymphoid cells (H9 or SupT1) were transfected with pNL4-3 or pNL4-3Δvif (2 μg), plus various plasmids (0.75 μg pcDNA3-CEM15 or its mutants), with FuGene 6 transfection reagents (Roche). The viruses in the supernatants were collected after 48 h. Of note, pNL4-3Δvif was constructed in our laboratory and has been described previously¹⁵.

Intracellular reverse transcription

Wild-type or Δvif viruses (10 ng of p24 equivalents), generated from either permissive cells or non-permissive cells, were treated with RQ1 DNase (5 U) (Promega) at 37 °C for 1 h and then allowed to infect C8166 T cells. Twelve hours after infection, the infected cells (in 150 μl TN buffer) were treated again with DNase (5 U) and incubated at 37 °C for 1 h. After washing, the newly synthesized viral DNA was extracted from the cells and amplified by PCR²⁰.

Endogenous reverse transcription

The wild-type or Δvif viruses (75 ng of p24 antigen equivalents) were concentrated by ultracentrifugation, followed by DNase treatment at 37 °C for 1 h. After washing off DNase, the viruses in the TN buffer were incubated with 1 mM dNTPs, 2.5 mM MgCl₂, and 15 μg ml⁻¹ melittin at 37 °C for 4 h^{20,29}. The newly synthesized DNA was isolated and amplified with PCR. The PCR products were inserted into the pGEM-T vector and individual clones were sequenced and analysed.

Viral passage experiments

The retroviral vectors (MLLV) carrying the CEM15 gene (pSLX-CMV-CEM15) or the vector alone were first transduced into SupT1 T cells, following methods described previously¹⁵. The naive C8166 or G418-resistant SupT1 T cells or C8166 cells were infected by wild-type or Δvif viruses (0.1 ng of p24 antigen equivalents), which were generated from 293T cells through transfection with pNL4-3 or pNL4-3Δvif. The infected T cells were then cultured in RPMI1640 medium plus 10% fetal bovine serum. After 2 weeks, the progeny viruses in the supernatants were allowed to infect C8166 or SupT1 cells, with or without CEM15, respectively. This procedure was repeated two or four times. The viral DNA in the final round of infected cells was extracted and amplified by PCR, followed by insertion into pGEM-T vector. The individual clones were then analysed by sequencing.

DNA PCR, RT-PCR and DNA/complementary DNA sequencing

The viral DNA was extracted from viruses or infected cells and PCR was performed, as described previously²⁰. The viral RNA was extracted from viruses or infected cells through RNeasy Mini Kits (Qiagen). DNase treatment was performed to eliminate any contaminating DNA. The RT-PCR was performed with ready to go RT-PCR beads (Amersham-Pharmacia). The PCR products were inserted into the pGEM-T vector and individual clones were sequenced. The mutation frequency (*f*) was calculated as: (number of substitutions)/(number of target nucleotides).

Cytidine deaminase assay

A method described previously³⁰ was performed to determine the cytidine deaminase activity of CEM15, with some modifications. Briefly, GST-CEM15 at various amounts was incubated with 4 μCi ³H-deoxycytidine (23 Ci mmol⁻¹; Moravsek) and 250 μM cytidine in a total volume of 10 μl Tris-HCl buffer (45 mM, pH 7.5). After incubation at 37 °C for 4–6 h, the reactions were quenched by adding 2 μl of 10 μg μl⁻¹ each of deoxycytidine and deoxyuridine. Four microlitres of sample was then added onto a silica-polyester TLC plate (Sigma-Aldrich). The plate was developed in 10:1:0.5 (v/v) ethyl acetate:methanol:trichloroacetic acid (6 M) for 2–3 h. The deoxycytidine and deoxyuridine bands were visualized by exposure to 254 nm ultraviolet light and cut separately into scintillation fluid for ³H-isotope quantifications.

Received 6 March; accepted 8 May 2003; doi:10.1038/nature01707.

Published online 28 May 2003.

1. Pathak, V. K. & Temin, H. M. Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc. Natl Acad. Sci. USA* **87**, 6019–6023 (1990).
2. Li, Y. *et al.* Molecular characterization of human immunodeficiency virus type 1 cloned directly from

uncultured human brain tissue: identification of replication-competent and -defective viral genomes. *J. Virol.* **65**, 3973–3985 (1991).

3. Vartanian, J. P., Meyerhans, A., Asjo, B. & Wain-Hobson, S. Selection, recombination, and G → A hypermutation of human immunodeficiency virus type 1 genomes. *J. Virol.* **65**, 1779–1788 (1991).
4. Vartanian, J. P., Meyerhans, A., Sala, M. & Wain-Hobson, S. G → A hypermutation of the human immunodeficiency virus type 1 genome: evidence for dCTP pool imbalance during reverse transcription. *Proc. Natl Acad. Sci. USA* **91**, 3092–3096 (1994).
5. Fitzgibbon, J. E., Mazar, S. & Dubin, D. T. A new type of G → A hypermutation affecting human immunodeficiency virus. *AIDS Res. Hum. Retroviruses* **9**, 833–838 (1993).
6. Vartanian, J. P., Henry, M. & Wain-Hobson, S. Sustained G → A hypermutation during reverse transcription of an entire human immunodeficiency virus type 1 strain Vau group O genome. *J. Gen. Virol.* **83**, 801–805 (2002).
7. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* **418**, 646–650 (2002).
8. Jarmuz, A. *et al.* An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285–296 (2002).
9. Strebel, K. *et al.* The HIV 'A' (sor) gene product is essential for virus infectivity. *Nature* **328**, 728–730 (1987).
10. Fisher, A. G. *et al.* The sor gene of HIV-1 is required for efficient virus transmission *in vitro*. *Science* **237**, 888–893 (1987).
11. Gabuzda, D. H. *et al.* Role of vif in replication of human immunodeficiency virus type 1 in CD4 + T lymphocytes. *J. Virol.* **66**, 6489–6495 (1992).
12. Sova, P. & Volsky, D. J. Efficiency of viral DNA synthesis during infection of permissive and nonpermissive cells with vif-negative human immunodeficiency virus type 1. *J. Virol.* **67**, 6322–6326 (1993).
13. von Schwedler, U., Song, J., Aiken, C. & Trono, D. Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. *J. Virol.* **67**, 4945–4955 (1993).
14. Simon, J. H. & Malim, M. H. The human immunodeficiency virus type 1 Vif protein modulates the postpenetration stability of viral nucleoprotein complexes. *J. Virol.* **70**, 5297–5305 (1996).
15. Zhang, H., Pomerantz, R. J., Dornadula, G. & Sun, Y. Human immunodeficiency virus type 1 Vif protein is an integral component of an mRNP complex of viral RNA and could be involved in the viral RNA folding and packaging process. *J. Virol.* **74**, 8252–8261 (2000).
16. Dettenhofer, M., Cen, S., Carlson, B. A., Kleiman, L. & Yu, X. F. Association of human immunodeficiency virus type 1 Vif with RNA and its role in reverse transcription. *J. Virol.* **74**, 8938–8945 (2000).
17. Khan, M. A. *et al.* Human immunodeficiency virus type 1 Vif protein is packaged into the nucleoprotein complex through an interaction with viral genomic RNA. *J. Virol.* **75**, 7252–7265 (2001).
18. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell* **10**, 1247–1253 (2002).
19. Petersen-Mahrt, S. K., Harris, R. S. & Neuberger, M. S. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99–103 (2002).
20. Dornadula, G., Yang, S., Pomerantz, R. J. & Zhang, H. Partial rescue of the vif-negative phenotype of mutant human immunodeficiency virus type 1 strains from nonpermissive cells by intravirion reverse transcription. *J. Virol.* **74**, 2594–2602 (2000).
21. Janini, M., Rogers, M., Birx, D. R. & McCutchan, F. E. Human immunodeficiency virus type 1 DNA sequences genetically damaged by hypermutation are often abundant in patient peripheral blood mononuclear cells and may be generated during near-simultaneous infection and activation of CD4(+) T cells. *J. Virol.* **75**, 7973–7986 (2001).
22. Kao, S. *et al.* Human immunodeficiency virus type 1 Vif is efficiently packaged into virions during productive but not chronic infection. *J. Virol.* **77**, 1131–1140 (2003).
23. Chester, A., Scott, J., Anant, S. & Navaratnam, N. RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA. *Biochim. Biophys. Acta* **1494**, 1–13 (2000).
24. Mehta, A., Banerjee, S. & Driscoll, D. M. Apobec-1 interacts with a 65-kDa complementing protein to edit apolipoprotein-B mRNA *in vitro*. *J. Biol. Chem.* **271**, 28294–28299 (1996).
25. Martinez, M. A., Vartanian, J. P. & Wain-Hobson, S. Hypermutagenesis of RNA using human immunodeficiency virus type 1 reverse transcriptase and biased dNTP concentrations. *Proc. Natl Acad. Sci. USA* **91**, 11787–11791 (1994).
26. Vartanian, J. P. *et al.* HIV genetic variation is directed and restricted by DNA precursor availability. *J. Mol. Biol.* **270**, 139–151 (1997).
27. Cattaneo, R. *et al.* Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell* **55**, 255–265 (1988).
28. Samuel, C. E. Antiviral actions of interferons. *Clin. Microbiol. Rev.* **14**, 778–809 (2001).
29. Zhang, H., Dornadula, G. & Pomerantz, R. J. Endogenous reverse transcription of human immunodeficiency virus type 1 in physiological microenvironments: an important stage for viral infection of nondividing cells. *J. Virol.* **70**, 2809–2824 (1996).
30. MacGinnitie, A. J., Anant, S. & Davidson, N. O. Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity. *J. Biol. Chem.* **270**, 14768–14775 (1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by research grants from the National Institute of Health (H.Z.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.Z. (hui.zhang@mail.tju.edu).

Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts

Bastien Mangeat^{*†}, Priscilla Turelli^{*†}, Gersende Caron[‡], Marc Friedli^{*§}, Luc Perrin[‡] & Didier Trono^{*§}

^{*} Department of Genetics and Microbiology, [‡] Department of Medicine, and [§] 'Frontiers in Genetics' Research Program, University of Geneva, 1211 Geneva 4, Switzerland

[†] These authors contributed equally to this work

Viral replication usually requires that innate intracellular lines of defence be overcome, a task usually accomplished by specialized viral gene products. The virion infectivity factor (Vif) protein of human immunodeficiency virus (HIV) is required during the late stages of viral production to counter the antiviral activity of APOBEC3G (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G; also known as CEM15), a protein expressed notably in human T lymphocytes^{1–4}. When produced in the presence of APOBEC3G, vif-defective virus is non-infectious. APOBEC3G is closely related to APOBEC1, the central component of an RNA-editing complex that deaminates a cytosine residue in *apoB* messenger RNA^{5–7}. APOBEC family members also have potent DNA mutator activity through dC deamination⁸; however, whether the editing potential of APOBEC3G has any relevance to HIV inhibition is unknown. Here, we demonstrate that it does, as APOBEC3G exerts its antiviral effect during reverse transcription to trigger G-to-A hypermutation in the nascent retroviral DNA. We also find that APOBEC3G can act on a broad range of retroviruses in addition to HIV, suggesting that hypermutation by editing is a general innate defence mechanism against this important group of pathogens.

As an initial step towards investigating the mechanism of APOBEC3G antiviral action, we asked whether the putative catalytic site of this protein is important for inhibiting vif-defective (Δ vif) HIV-1. For this, we used single-round assays to measure the infectivity of wild-type and Δ vif particles produced by transient transfection of 293T cells, in the presence or absence of this cellular factor (Fig. 1). APOBEC3G, which localized to the cytoplasm of the transfected cells (Fig. 1a), induced, as previously observed⁴, a marked decrease in the infectivity of Δ vif HIV-1 (Fig. 1b). We then tested a series of APOBEC3G point mutants, concentrating on residues found in the zinc-coordinating catalytic site of APOBEC family members^{5,9}. Mutating critical glutamate and cysteine residues (E67Q, C100S, E259Q and C291S)—previously shown to be essential for APOBEC1 function—abrogated the inhibitory potential of APOBEC3G, whereas changing these amino acids at several other positions (E85Q, C221S, E323Q) was without consequence (Fig. 1c).

These results strongly suggested that APOBEC3G inhibits retroviral infectivity by means of editing. However, the target of this process remained undefined. We thus examined the breadth of APOBEC3G antiviral activity (Fig. 2a). Using HIV-derived lentiviral vector particles produced from a multiply-deleted packaging construct lacking all the HIV-1 accessory genes (*vif*, *vpr*, *vpu* and *nef*) and pseudotyped with the vesicular stomatitis virus G (VSV-G) envelope protein¹⁰, we found that APOBEC3G induced a Vif-responsive inhibition of transduction. APOBEC3G action and its blockade by Vif are thus independent of the route of viral entry and from these other viral gene products. Because editing enzymes often act in a sequence-specific manner, we asked whether APOBEC3G could also inhibit HIV-derived vectors produced from two 'codon-optimized' packaging constructs. The *gag* and *pol* genes of these

constructs contain close to 1,000 silent nucleotide substitutions compared with the original HIV-derived lentivector system, but they produce virions with the same protein composition¹¹. In spite of these sequence differences, APOBEC3G was fully active on these viruses. Furthermore, it also inhibited the infectivity of retroviral particles only distantly related to HIV-1, such as vectors derived from simian immunodeficiency virus (SIV)¹², equine infectious anaemia virus (EIAV)¹³ and murine leukaemia virus (MLV). In all

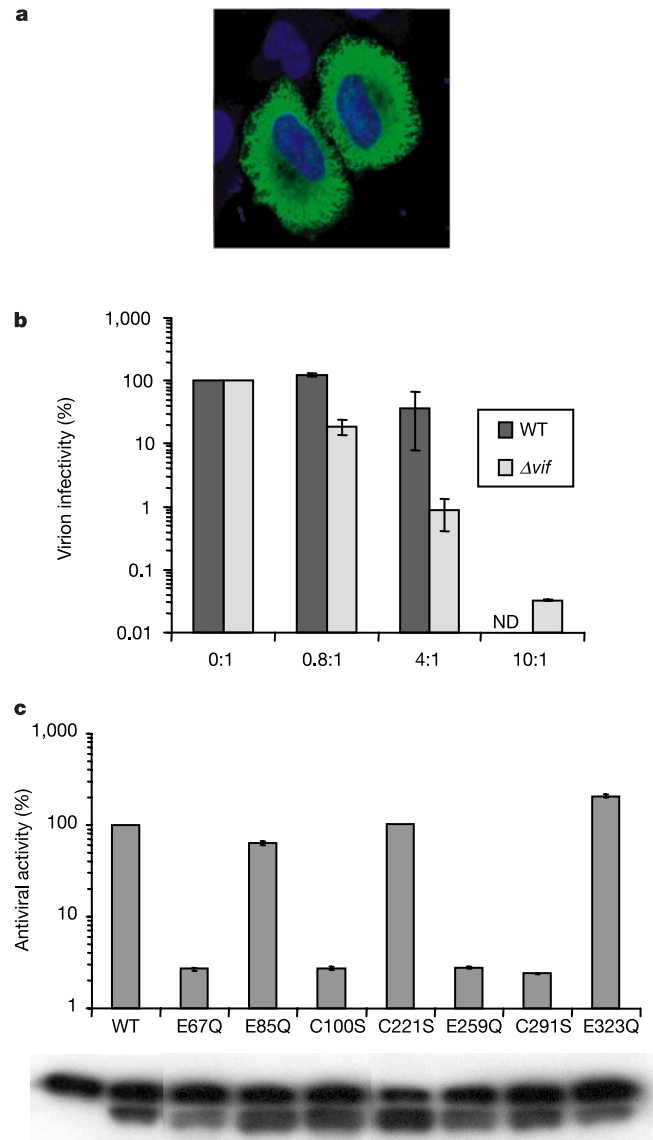


Figure 1 The putative catalytic site of APOBEC3G is important for HIV inhibition.

a, Subcellular localization of APOBEC3G in transfected HeLa cells. The HA tag of the protein served for its detection by indirect immunofluorescence. 293T cells used for viral production gave the same image. **b**, Dose-dependent inhibition of Δ vif HIV-1 infectivity by APOBEC3G, determined by single-round titration in CD4⁺ P4 HeLa cells. The molar ratios of APOBEC3G compared with proviral DNA plasmids used for transfection are indicated on the x axis. Titres obtained with each virus produced in the absence of APOBEC3G were given the arbitrary value of 100%. ND, not determined; WT, wild type. **c**, Testing of indicated APOBEC3G point mutants, performed as in **b** using a 5:1 molar ratio of APOBEC3G/ Δ vif proviral DNA plasmids. Below is a western blot analysis of extracts from virus-producing 293T cells, using antibodies against tubulin (top band) and the HA tag of APOBEC3G (bottom band). The lane on the far left corresponds to untransfected 293T cells.

of these cases, inhibition was prevented by HIV-1 Vif, even for vectors derived from EIAV and MLV, two viruses devoid of the *vif* gene. Vif can thus act independently of other HIV-specific factors. Notably, APOBEC3G was detected in MLV-derived particles at levels roughly comparable to those found in HIV-derived virions (Fig. 2b).

The broad spectrum of APOBEC3G action suggested at least two possibilities. First, it might modify the sequence coding for a cellular cofactor essential for the generation of infectious retroviruses. Second, it might act on some viral genetic material, but in that case not in a sequence-specific manner considering the low level of identity between the genomes of HIV-1, SIV, EIAV and MLV. Although either hypothesis seemed plausible, the second one received support from a re-examination of the phenotypic consequences of APOBEC3G action.

We first noted that the infectivity defect of Δvif HIV-1 produced by proviral DNA transfection of H9 T lymphoid cells (which naturally express APOBEC3G) was only 50-fold at most, whereas this virus was completely unable to spread in this cell type, as if the

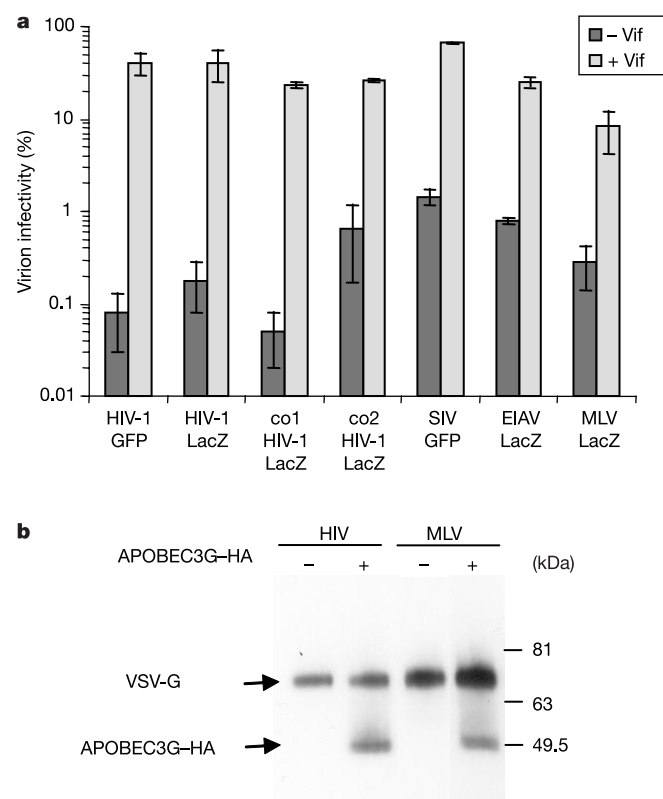


Figure 2 APOBEC3G is a broadly active antiretroviral. **a**, Single-round measurement of APOBEC3G-mediated inhibition of transduction of HeLa cells with particles derived from the following retroviral vector systems: multiply-deleted HIV-1-derived lentiviral vectors (HIV-1-GFP and HIV-1-LacZ); two 'codon-optimized' derivatives (co1 HIV-1-LacZ and co2 HIV-1-LacZ); non-human lentiviruses SIV (SIV-GFP) and EIAV (EIAV-LacZ); oncoretrovirus MLV (MLV-LacZ). Transduction efficiencies were scored by FACS (for GFP-expressing vectors) or 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal) staining (for LacZ-expressing vectors). Transduction efficiency of vector produced in the absence of APOBEC3G was given the arbitrary value of 100% for each vector. Where indicated, HIV-1 Vif was expressed in vector producer cells. **b**, APOBEC3G is incorporated into MLV particles. VSV-G-pseudotyped HIV-1 and MLV particles were produced from transiently transfected 293T cells in the presence or absence of HA-tagged APOBEC3G. Normalized amounts (as assessed by reverse transcription activity) of Optiprep gradient-purified virions were analysed by western blotting, using antibodies against HA and VSV-G. Of note, the normalization is approximate, as it relies on measuring the activity of two distinct reverse transcriptases.

magnitude of its replication defect increased with additional rounds of infection (data not shown). To probe this issue further, we infected T lymphoid SupT1 cells with VSV-G-pseudotyped wild-type and Δvif HIV-1 released from control or APOBEC3G-expressing 293T cells. SupT1 cells do not express APOBEC3G and are thus

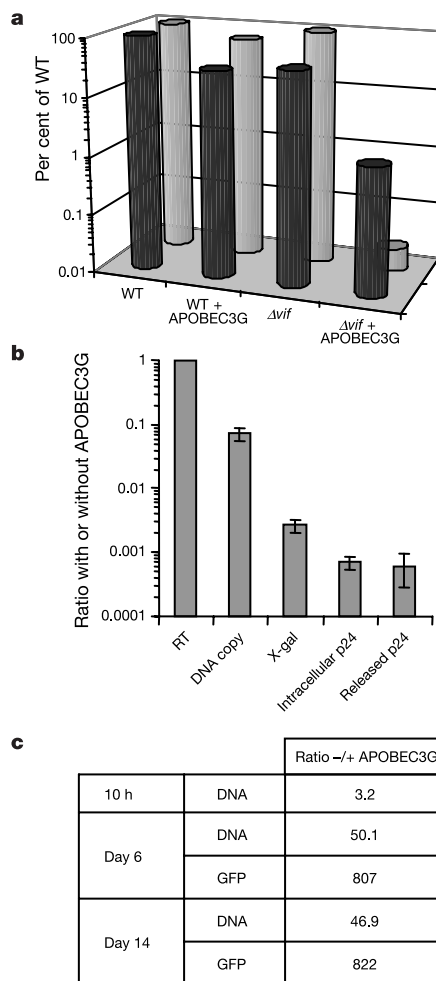
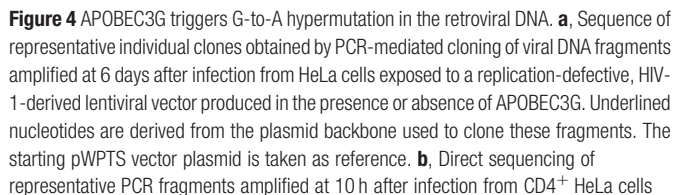


Figure 3 Decreased viral DNA yields do not fully account for APOBEC3G-mediated inhibition. **a**, SupT1 human T lymphoid cells were infected with VSV-G-pseudotyped wild-type or Δvif HIV-1, produced in the presence or absence of APOBEC3G from transiently transfected 293T cells. After overnight incubation, the input virus was removed by extensive washing, and 60 h later the viral DNA content of the lymphoid cells was determined by real-time PCR, and their infectious virus output measured by titration on CD4⁺ HeLa-derived P4 cells. The absence of significant amounts of the original 293T-produced VSV-G-pseudotyped virus in these supernatants was verified by parallel titration on CD4⁺ HeLa cells (not shown). Black bars, viral DNA content of SupT1 cells/input virus (reverse transcription activity); grey bars, infectious output/viral DNA content of SupT1 cells. Numbers obtained for wild-type virus produced in the absence of APOBEC3G were given the arbitrary score of 100%. Results are representative of a duplicate experiment. **b**, CD4⁺, HIV-1 LTR- β -gal-containing HeLa-derived P4 cells were infected as indicated above with Δvif HIV-1 exposed or not to APOBEC3G during production. After 24 h, target cells were extensively washed and 50 μ M AZT was added. Sixty hours later, these cells were analysed for viral DNA load (by real-time PCR), Tat-induced β -gal expression (by X-gal staining), intracellular p24 content (by FACS) and virion release (by p24 ELISA on supernatants). Values were normalized for input virus (as determined by reverse transcription activity). **c**, HeLa cells were infected with a GFP-expressing, HIV-1-derived lentiviral vector produced without Vif in the presence or absence of APOBEC3G. Viral DNA content and rates of GFP expression of the target cells were evaluated at indicated times by real-time PCR and FACS analysis, respectively. Results are expressed as the ratio of values obtained from virus produced in the absence versus the presence of APOBEC3G.

To investigate this issue further, we examined the viral DNA load and the transgene expression profile of HeLa cells exposed to a replication-defective, green fluorescent protein (GFP)-expressing, HIV-1-derived lentiviral vector made from a *vif*-less packaging construct (Fig. 3c). At 10 h after infection, levels of reverse transcripts induced by this vector were decreased threefold if it had been produced in the presence of APOBEC3G. At six days, this difference had increased to 50-fold. Viral DNA levels in the target cell population then remained constant, as assessed by a third measurement performed at 14 days after infection. Because these cells were rapidly dividing, the viral DNA detected from the sixth day on was stably integrated. The increasing viral DNA defect noted at 6 and 14 days compared with 10 h after infection was consistent with a destabilizing effect of APOBEC3G on the HIV-1 reverse transcription complex, which not only compromised the synthesis of the viral DNA but triggered its degradation before it could integrate, as previously described for Δvif HIV-1 produced from restrictive cells^{1,3}. An additional and crucial piece of information was obtained by comparing the integrated viral DNA content with the apparent rate of transduction of the target cells. The 50-fold lower proviral DNA load of cells exposed to the vector produced in the presence of APOBEC3G was indeed markedly outmatched by their 800-fold



NATURE | VOL 424 | 3 JULY 2003 | www.nature.com/nature

lower rate of GFP positivity, compared with cells infected with the control vector. Therefore, only one in every 16 transduced cells appeared capable of expressing the vector transgene. A similar type of result was obtained with MLV-derived vectors (data not shown).

This accumulation of non-functional proviruses prompted us to sequence viral DNAs in the target cells. This revealed that *vif*-less vector produced in the presence of APOBEC3G yielded DNA with a very high number of G-to-A substitutions (Fig. 4a). These were noted with equal frequency in DNA isolated at 10 h, 6 days and 14 days after infection (data not shown). Altogether, approximately 13% of G residues were mutated to A in the regions examined. The likelihood of this nucleotide change was higher if the original sequence contained stretches of two or more G residues (4% compared with 32% chance of G-to-A mutation for single versus multiple G residues). Importantly, no other type of nucleotide substitution was found in more than 10,000 sequenced bases. A similarly high rate of G-to-A mutation was detected in proviral DNA derived from Δvif HIV-1 virions (Fig. 4b) and from MLV retroviral vector particles (data not shown) produced from APOBEC3G-expressing cells. Correlating its ability to counter the inhibitory effect of APOBEC3G on HIV-1 replication, *Vif* expression in virus producer cells prevented the accumulation of G-to-A mutations (data not shown).

These results are consistent with a model in which the previously demonstrated cytidine deaminase activity of APOBEC3G leads to the replacement of cytosine by uracil in the nascent retroviral DNA. Indeed, if the viral RNA were the target of this modification, C-to-T rather than G-to-A changes would be observed in the viral DNA. Confirming this point, partial sequencing of the genomic RNA of APOBEC3G-exposed Δvif HIV-1 virions did not reveal any mutation (data not shown). The increased frequency of G-to-A changes in poly-G stretches indicates that cytosine residues are targeted by APOBEC3G in the context of the DNA, and not as free nucleotides. It has already been noted that APOBEC3G favours certain hotspots as substrates for cytidine deamination⁸.

The replacement of cytosine by uracil in the minus strand DNA during reverse transcription leads to G-to-A transitions in the plus strand. In contrast, a C-to-U change in the plus strand would probably be corrected using the minus strand as a template. Therefore, we cannot exclude the possibility that APOBEC3G affects both DNA strands. However, it is interesting to note that the related enzyme activation-induced cytidine deaminase (AID) can deaminate cytidines on single-stranded but not double-stranded DNA¹⁵.

The incorporation of uracil into minus strand DNA was recently shown to interfere with initiation of plus strand synthesis during lentiviral reverse transcription¹⁶. This may contribute to the decreased amounts of viral DNA found in Δvif -infected cells. Although our results do not directly explain why APOBEC3G also leads to the degradation of synthesized yet unintegrated reverse transcripts, this latter phenomenon may be one consequence of the handling of uridine-rich DNA by repair pathways. In this respect, HIV-1 packages the uracil DNA glycosylase hUNG-2, a host-derived enzyme that removes uracil bases from DNA as part of the base excision repair pathway. Although it may correct some of the G-to-A changes induced by APOBEC3G, it remains that HIV-1 proviruses established under the influence of the antiviral are highly mutated and as a consequence are largely non-functional.

The relative lack of sequence specificity of APOBEC3G-mediated editing explains the broad spectrum of activity of this antiviral factor, which can both inhibit a wide range of retroviruses and compromise expression from very different proviruses. The APOBEC3G-sensitive MLV and ELAV do not encode a *vif* gene. Perhaps a *vif*-like activity buried in some other gene product allows them to counter the orthologue of APOBEC3G probably present in their cognate host. Alternatively, they may preferentially replicate in APOBEC3G-negative cells, as this gene exhibits a narrowly restricted expression pattern, at least in humans⁴.

G-to-A hypermutation was previously observed in HIV and other lentiviruses^{17–21}, and to a lesser extent in other retroviruses such as human T-cell leukaemia virus²², spleen necrosis virus²³, a close relative of MLV, and in the pararetrovirus hepatitis B virus²⁴. Whether APOBEC3G-mediated editing has a role in these cases as well will need to be explored. □

Methods

Constructs, virus productions, infections and titrations

Wild-type and *vif*-defective HIV-1 proviral clones, as well as the HIV-derived lentiviral vector system, have been described previously^{1–10}. For the latter, packaging and envelope functions were provided by plasmids pCMV- $\Delta R8.91$ and pMD.G, respectively, whereas the transfer vector was the GFP-containing pWPTS plasmid (see <http://www.tronolab.unige.ch> for details). The codon-optimized HIV vector packaging systems pHDMHgp2 and pSYNGP¹¹ were obtained from R. Mulligan and K. Mitrophanous, respectively. The SIV (obtained from F.-L. Cosset) and ELAV vector systems were as described^{12,13}. The plasmid expressing a haemagglutinin (HA)-tagged form of APOBEC3G⁴ was a gift from M. Malim. APOBEC3G mutants were constructed with the Quickchange Mutagenesis kit (Stratagene). HIV-1 and retroviral vector particles were produced by transient transfection of 293T cells with Fugene (Roche) (see <http://www.tronolab.unige.ch> for details). We performed infections and titrations as previously described²⁵.

Protein analyses

HA-tagged APOBEC3G indirect immunofluorescence was performed as previously described²⁵, using the monoclonal HA.11 antibody (BabCO). For HA- or VSV-G-specific western blot analyses, the mouse monoclonal 3F10-peroxidase-conjugated antibody (Roche) and a rabbit polyclonal antibody (a gift from M. Matsuda) were used, respectively. Virions were isolated by ultracentrifugation through a sucrose cushion followed by purification through an Optiprep (Axis-Shield) gradient²⁶, to minimize contamination by nonspecific microvesicles. Intracellular p24 fluorescence-activated cell sorting (FACS) analysis was performed with the Cytotix/Cytoperm kit (BD Biosciences).

Real-time PCR measurement of viral DNA load

Cells were infected with DNase-treated viruses or vectors and DNA was extracted with DNeasy kit (Qiagen). Taqman quantitative polymerase chain reaction (PCR) (Applied Biosystem) of virus sequences was performed with late reverse transcripts primers and probe²⁷. Vector sequences were detected with GFP primers (GFP sense, 5'-CTGCTGCCCCACAACCAC-3'; GFP antisense, 5'-ACCATGTGATCGCGCTTCTC-3'; GFP probe, FAM-CCAGTCCGCCCTGAGCAAGACC-TAMRA). The specific amplification of neo-synthesized reverse transcripts was controlled by AZT (zidovudine) treatment of a subset of target cells.

Sequencing

DNA was obtained as for Taqman PCR. Limited dilutions were either submitted to amplification (high fidelity Pfu Turbo polymerase; Stratagene) with primers specific for 5' long terminal repeat (LTR) (PB.nef-1 sense, 5'-AGGAGCTGTAGATCTTAGCCACTT-3'; PB.U5-1 antisense, 5'-GGTCTGAGGATCTCTAGTTAC-3') and directly sequenced²⁸, or submitted to amplification with the same primers flanked by 5' *Bam*HI and 3' *Eco*RI sites, cloned and sequenced. RNA from DNase-treated virions used for the infections was sequenced via the PCR-mediated amplification of complementary DNA synthesized with AMV reverse transcriptase.

Received 14 April; accepted 13 May 2003; doi:10.1038/nature01709.

Published online 28 May 2003.

1. von Schwedler, U., Song, J., Aiken, C. & Trono, D. *Vif* is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. *J. Virol.* **67**, 4945–4955 (1993).
2. Gabuzda, D. H. *et al.* Role of *vif* in replication of human immunodeficiency virus type 1 in CD4⁺ T lymphocytes. *J. Virol.* **66**, 6489–6495 (1992).
3. Simon, J. H. & Malim, M. H. The human immunodeficiency virus type 1 *Vif* protein modulates the postpenetration stability of viral nucleoprotein complexes. *J. Virol.* **70**, 5297–5305 (1996).
4. Sheehy, A. M., Gaddis, N. C., Choi, J. D. & Malim, M. H. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral *Vif* protein. *Nature* **418**, 646–650 (2002).
5. Jarmuz, A. *et al.* An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* **79**, 285–296 (2002).
6. Powell, L. M. *et al.* A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* **50**, 831–840 (1987).
7. Teng, B., Burant, C. F. & Davidson, N. O. Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* **260**, 1816–1819 (1993).
8. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell* **10**, 1247–1253 (2002).
9. MacGinnitie, A. J., Anant, S. & Davidson, N. O. Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity. *J. Biol. Chem.* **270**, 14768–14775 (1995).
10. Zufferey, R., Nagy, D., Mandel, R. J., Naldini, L. & Trono, D. Multiply attenuated lentiviral vector achieves efficient gene delivery *in vivo*. *Nature Biotechnol.* **15**, 871–875 (1997).
11. Kotsopoulou, E., Kim, V. N., Kingsman, A. J., Kingsman, S. M. & Mitrophanous, K. A. A Rev-independent human immunodeficiency virus type 1 (HIV-1)-based vector that exploits a codon-optimized HIV-1 gag-pol gene. *J. Virol.* **74**, 4839–4852 (2000).
12. Negre, D. *et al.* Characterization of novel safe lentiviral vectors derived from simian

- immunodeficiency virus (SIVmac251) that efficiently transduce mature human dendritic cells. *Gene Ther.* **7**, 1613–1623 (2000).
13. Mitrophanous, K. *et al.* Stable gene transfer to the nervous system using a non-primate lentiviral vector. *Gene Ther.* **6**, 1808–1818 (1999).
 14. Charneau, P. *et al.* HIV-1 reverse transcription. A termination step at the center of the genome. *J. Mol. Biol.* **241**, 651–662 (1994).
 15. Jayanta, C., Ming, T., Chan, K., Eric, P. & Alt, F. W. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **422**, 726–730 (2003).
 16. Klarman, G. J., Chen, X., North, T. W. & Preston, B. D. Incorporation of uracil into minus strand DNA affects the specificity of plus strand synthesis initiation during lentiviral reverse transcription. *J. Biol. Chem.* **278**, 7902–7909 (2003).
 17. Borman, A. M., Quillent, C., Charneau, P., Kean, K. M. & Clavel, F. A highly defective HIV-1 group O provirus: evidence for the role of local sequence determinants in G → A hypermutation during negative-strand viral DNA synthesis. *Virology* **208**, 601–609 (1995).
 18. Vartanian, J. P., Meyerhans, A., Asjo, B. & Wain-Hobson, S. Selection, recombination, and G → A hypermutation of human immunodeficiency virus type 1 genomes. *J. Virol.* **65**, 1779–1788 (1991).
 19. Wain-Hobson, S., Sonigo, P., Guyader, M., Gazit, A. & Henry, M. Erratic G → A hypermutation within a complete caprine arthritis-encephalitis virus (CAEV) provirus. *Virology* **209**, 297–303 (1995).
 20. Johnson, P. R., Hamm, T. E., Goldstein, S., Kitov, S. & Hirsch, V. M. The genetic fate of molecularly cloned simian immunodeficiency virus in experimentally infected macaques. *Virology* **185**, 217–228 (1991).
 21. Turelli, P., Guigney, F., Mornex, J. F., Vigne, R. & Querat, G. dUTPase-minus caprine arthritis-encephalitis virus is attenuated for pathogenesis and accumulates G-to-A substitutions. *J. Virol.* **71**, 4522–4530 (1997).
 22. Mansky, L. M. *In vivo* analysis of human T-cell leukemia virus type 1 reverse transcription accuracy. *J. Virol.* **74**, 9525–9531 (2000).
 23. Pathak, V. K. & Temin, H. M. Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc. Natl Acad. Sci. USA* **87**, 6019–6023 (1990).
 24. Gunther, S. *et al.* Naturally occurring hepatitis B virus genomes bearing the hallmarks of retroviral G → A hypermutation. *Virology* **235**, 104–108 (1997).
 25. Turelli, P. *et al.* Cytoplasmic recruitment of IN1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication. *Mol. Cell* **7**, 1245–1254 (2001).
 26. Dettenhofer, M. & Yu, X. F. Highly purified human immunodeficiency virus type 1 reveals a virtual absence of Vif in virions. *J. Virol.* **73**, 1460–1467 (1999).
 27. Butler, S. L., Hansen, M. S. & Bushman, F. D. A quantitative assay for HIV DNA integration *in vivo*. *Nature Med.* **7**, 631–634 (2001).
 28. Yerly, S. *et al.* Transmission of antiretroviral-resistant HIV-1 variants. *Lancet* **354**, 729–733 (1999).

Acknowledgements We thank S. Vianin for technical assistance, and M. Malim, R. Mulligan, K. Mitrophanous, F.-L. Cosset and M. Matsuda for the gift of reagents. This work was supported by the Swiss National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.T. (didier.trono@medecine.unige.ch).

Processive AID-catalysed cytosine deamination on single-stranded DNA simulates somatic hypermutation

Phuong Pham, Ronda Bransteitter, John Petruska & Myron F. Goodman

Departments of Biological Sciences and Chemistry, Hedco Molecular Biology Laboratories, University of Southern California, University Park, Los Angeles, California 90089-1340, USA

Activation-induced cytidine deaminase (AID) is a protein required for B cells to undergo class switch recombination and somatic hypermutation (SHM)—two processes essential for producing high-affinity antibodies¹. Purified AID catalyses the deamination of C to U on single-stranded (ss)DNA^{2–4}. Here, we show *in vitro* that AID-catalysed C deaminations occur preferentially on 5' WRC sequences in accord with SHM spectra observed *in vivo*. Although about 98% of DNA clones suffer no mutations, most of the remaining mutated clones have 10–70 C to T transitions per clone. Therefore, AID carries out multiple C

deaminations on individual DNA strands, rather than jumping from one strand to another. The avid binding of AID to ssDNA could result from its large net positive charge (+11) at pH 7.0, owing to a basic amino-terminal domain enriched in arginine and lysine. Furthermore, AID exhibits a 15-fold preference for C deamination on the non-transcribed DNA strand exposed by RNA polymerase than the transcribed strand protected as a RNA–DNA hybrid. These deamination results on ssDNA bear relevance to three characteristic features of SHM: preferential mutation at C sites within WRC hotspot sequences, the broad clonal mutagenic heterogeneity of antibody variable regions targeted for mutation^{5,6}, and the requirement for active transcription to obtain mutagenesis^{7,8}.

AID was identified three years ago^{9,10}, and knowledge of its function in the generation of high-affinity antibodies has helped to reveal why children lacking AID exhibit an inefficient antibody response¹⁰. Despite the similarity of AID in sequence to APOBEC1, a cytidine deaminase involved in messenger RNA editing¹¹, initial AID purifications indicated no discernible activity on nucleic acid substrates *in vitro*¹². Recently, however, we were able to show that human AID purified from *Baculovirus*-infected insect cells catalyses the deamination of C → U on ssDNA², but not on double-stranded (ds)DNA, RNA–DNA hybrids or ssRNA², with similar conclusions drawn by other groups using *Escherichia coli*⁴ and activated B-cell extracts³ to purify AID. SHM-like spectra were observed in *E. coli*, DT40 chicken B cells and in mice lacking Ung activity expressing endogenous levels of AID^{6,13,14}, demonstrating that AID-catalysed C deamination has a definitive role in the mutational targeting process. In an effort to reconstitute SHM *in vitro*, we have examined the action of AID on a long ssDNA target sequence including a *lacZ* alpha (*lacZa*) reporter gene. We find that the action of AID yields a mutation spectrum with hotspots and cold spots much like those observed for SHM *in vivo*. The action of AID also shows a highly diverse clonal pattern of mutations, leaving many ssDNA molecules unmutated while introducing multiple mutations into a few molecules (Fig. 1).

AID was incubated with phage M13mp2 circular DNA substrate containing a 230-nucleotide target of the *lacZa* reporter sequence in a 365-nucleotide, single-stranded gapped region (Fig. 1a). This system has been used extensively to measure the fidelity of DNA polymerases by scoring mutated (white and light-blue plaques) and non-mutated (dark-blue plaques) phage progeny after transfection into *E. coli*¹⁵. AID-catalysed deamination events occurring on individual ssDNA *lacZa* reporter sequences are detected as C → T

Table 1 AID-catalysed C → U sequences

Hotspots		Cold spots	
Site	No. of C → U	Site	No. of C → U
ATACGC	34	TGCCCG	0
ATGCTT	34	TCCCGA	0
CAGCTA	33	CGCCTT	0
CAACTT	32	CGCCAG	0
ATGCAG	31	GGCCGA	0
AAACCC	31	TTCAC	1
TTACCC	31	CGGCTC	1
AAAGCA	29	TCAAC	1
TAGCTG	28	ACAAG	1
AcGCAA	28	GGCCGT	1
AAACCG	28	CGTCTG	2
TTGCAG	28	TaaCAA	2
CAGCAC	28	TGgCCG	3
AAACAG	26	CGCCCA	3
TTACGA	26	TCTCAC	3
		CGaCAG	3
		CGTCTG	4
WRC		SYC	

Hotspots and cold spots were derived from the AID-catalysed deamination spectrum shown in Fig. 2. Hotspots are defined as C deamination sites (underlined) occurring in ≥50% of the mutant phage. Cold spots are C sites occurring in ≤8% of the mutated phage. Mismatches within the consensus motifs are indicated by lower-case letters. All sequences are indicated 5' to 3'. The consensus motifs are indicated in bold. W = A or T; S = G or C; R = purine; Y = pyrimidine.

mutations in individual phage progeny after transfection into uracil glycosylase-deficient (Ung^-) *E. coli*. We measured a 32-fold increase in mutations comparing AID-treated DNA transfected in Ung^+ and Ung^- *E. coli* backgrounds (data not shown). AID is solely responsible for generating the C deaminations leading to C $\rightarrow$ T transitions because the mutations occurring at the deaminated C sites are not observed above background, either when transfection takes place in Ung^+ cells or when AID-treated DNA is incubated in the presence of purified uracil glycosylase before transfection into Ung^- cells (data not shown). In Ung^+ cells, uracil residues are removed by endogenous glycosylase followed by cleavage of the abasic sites by AP endonuclease, resulting in biologically inactive M13 DNA.

Two principal features of AID activity observed here *in vitro* that relate to the distinctive SHM spectra *in vivo* are the presence of a heterogeneous distribution of mutation frequencies in individual clones (Fig. 1b, c) and the sequence context-dependence of AID activity mimicking SHM hotspots (Table 1 and Fig. 2). A broad distribution of C $\rightarrow$ T mutation frequencies is observed for individually sequenced 'mutant' DNA molecules: about half having 1–20 mutations per clone and half with 21–80 mutations per clone (Fig. 1b).

Despite the occurrence of multiple mutations on a single DNA strand, the mutated targets comprise only about 2% of total target DNA. The absence of mutations on 98% of the single-stranded target DNA was determined by direct sequencing of the entire gapped DNA region purified from 40 randomly selected dark-blue (wild type) clones. We also found no mutations when sequencing the dsDNA region, thereby confirming our previous observation

that AID is active only on ssDNA². We have verified, moreover, that a mutant AID (H56R/E58Q) fails to act on ssDNA (data not shown).

The absence of any measurable deamination on 98% of the target ssDNA shows that the deamination events are not distributive, indicating that AID may act processively. Once having bound to a ssDNA substrate, AID remains bound long enough to catalyse multiple deaminations. These data alone do not rule out that a bound AID might aggregate with other AID molecules; however, our experiments with the gapped M13 DNA substrate have about equal numbers of unannealed 7.2-kilobase (kb) ssDNA and gapped circular dsDNA substrate, with 365 nucleotides of ssDNA in the gap. Assuming that AID binds roughly in proportion to ssDNA length, then about 1 out of 20 (that is, $365/(7,200 + 365)$) AID molecules would be available to bind to the 365-nucleotide gap. Therefore, aggregation of AID molecules within the gap seems unlikely. More specifically, we also observe clones containing as many as five consecutive deaminated C residues (for example, Fig. 1c clone 9). This observation is unlikely to arise either from the action of a single immobile AID molecule—because a 25-kDa protein probably has a footprint exceeding one DNA nucleotide—or from a larger immobile aggregate of AID molecules. More generally, 80% of the clones contain different combinations of C $\rightarrow$ T mutations at CC and CNC sequences. These 'close-packed' deaminations are simplest to explain by having a mobile AID molecule acting to deaminate C residues in a sequence-dependent manner. Therefore, whether or not AID aggregates, deaminations seem to be clustered in a way suggesting, although not proving, that AID is acting

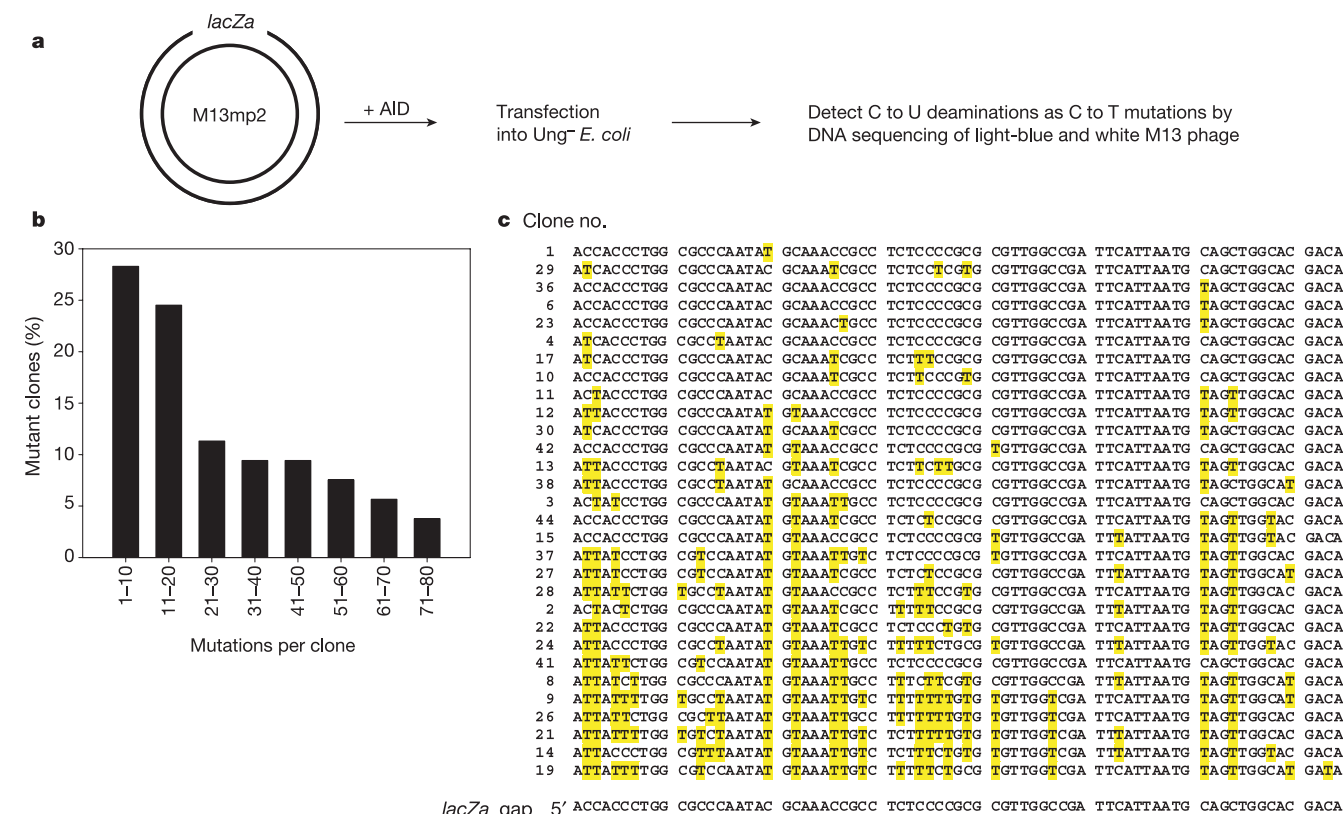


Figure 1 Processive AID-catalysed C deamination on ssDNA recapitulates SHM hotspots and clonal mutational heterogeneity. **a**, Sketch of protocol to detect C $\rightarrow$ T mutations resulting from AID-catalysed C deaminations occurring on individual ssDNA substrate molecules in a gapped M13 DNA construct. The ssDNA gap region is 365 nucleotides long and contains 230 nucleotides of *lacZa* reporter sequence. **b**, The distribution of mutant

clones exhibiting various numbers of C $\rightarrow$ T mutations. **c**, DNA sequencing data obtained from a subset of 30 individual M13 mutant DNA clones, showing the sequence context surrounding C $\rightarrow$ T mutations (highlighted in yellow) in the first 74 nucleotides of the 365-nucleotide gapped ssDNA region.

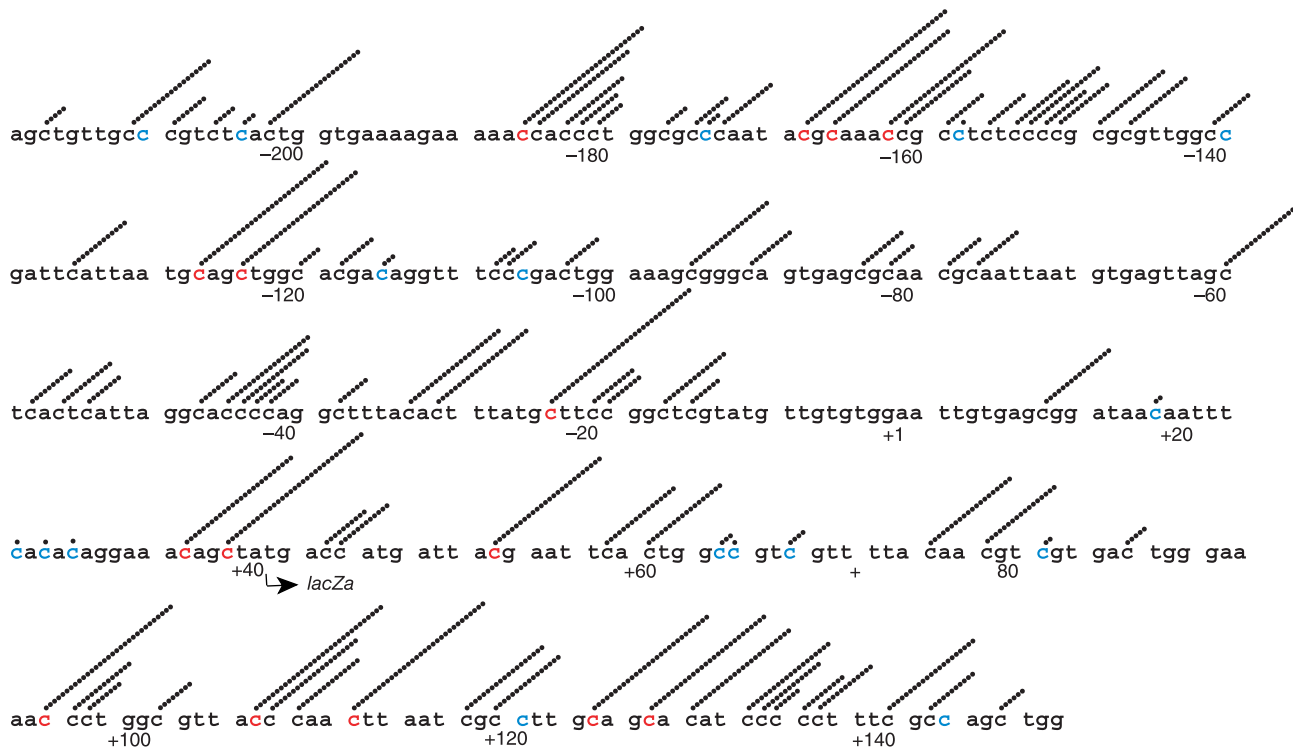


Figure 2 AID-catalysed C → U deamination spectrum. Each dot denotes a single C deamination taking place within the 365-nucleotide gapped M13 ssDNA region. Deaminations are identified as C → T mutations occurring in M13 DNA isolated from either white or light-blue plaques (see Methods). The hotspot C sites listed in Table 2 are

shown in red; cold spots are shown in blue. A total of 1,381 mutations at 106 C sites are shown. Not shown are two clones without C → T mutations that probably arose from pre-existing spontaneous mutations within the DNA gap; one has a C addition at position +132 in *lacZa* and the other contains a C → G mutation at position +129.

processively. Further biochemical analysis will be required to pin down the precise mechanism.

The presence of a large net positive charge (+11 at pH 7.0), owing to a basic N-terminal domain rich in Arg and Lys (Swiss-Prot and TrEMBL databases, March 2003), might enable AID to bind single-stranded nucleic acid (ssDNA) as its natural substrate, and ssRNA as a tightly bound inhibitor of AID². APOBEC1, the only other C deaminase known to act on a nucleic acid substrate (mRNA), also bears a positive charge (+9 at pH 7.0). In contrast, APOBEC2 and -3, as well as all known scavenging mononucleoside and mononucleotide C deaminases, have a net charge close to zero or are negatively charged.

A major conclusion derived from the biochemical data is that

AID deamination activity varies significantly with sequence context. The deamination of C is clearly favoured in WRC (where R = purine, W = A or T) contexts (Tables 1 and 2), reflecting variable (V)-gene WRC (or complementary GYW) mutational hotspots^{16–19}. The strong dependence of AID activity on sequence context is revealed by comparing the 15 hottest hexanucleotide regions, each hit by deamination 26–34 times, to the 17 coldest hexanucleotide regions, each with only 0–4 deamination hits (Table 1 and Fig. 2). Notably, 14 of the 15 hottest sequences have the WRC hotspot motif, the sole exception being a CGC sequence (ACGCAA), which was hit 28 times (Table 1).

The average number of deamination events per C site *in vitro* irrespective of sequence context is 13 (1,381 mutations scored/106 distinct sites; see Supplementary Information). The bases present at the –1 and –2 positions appear to exert the most significant influence on AID activity ($P < 0.001$). The observed number of C deamination events is 1.5- to 2-fold higher when a purine is the 5' nearest neighbour of C (that is, –1 position); there is a 1.7- to 4-fold increase when A or T is in the –2 position. Although the hotspot motif most typically referred to in SHM analysis is WRCY¹⁶, a preference for pyrimidine (Y) at the +1 position is not observed in our data (Table 1). We note that Y at the +1 site in the immunoglobulin V gene is over-represented owing to the absence of G caused by methyl-directed mutations at CG sites in vertebrate DNA, and hence may not be a bona fide contributor to a hotspot consensus sequence for SHM.

A mutability index parameter, defined as the frequency of observed C target base mutations compared to the expected unbiased mutation frequency²⁰, is useful for making a more quantitative comparison (Table 2). For AID-catalysed C → U, we find mutability index values between 1.45 and 2.15, averaging 1.85 ± 0.26 in WRC hotspots (Table 2), but only 0.05–0.64, averaging 0.40 ± 0.25 , in SYC cold spots. These results compare

Table 2 Comparison of mutability index for sequences for AID deamination and SHM

Sequence (5' to 3')	AID <i>in vitro</i>	SHM <i>in vivo</i>
Hotspots		
AAC	1.72	2.00 (3.41)
AGC	1.45	2.58 (3.38)
TAC	2.07	2.24 (2.89)
TGC	2.15	0.71 (2.09)
WRC average	1.85 ± 0.26	1.88 ± 0.59
Cold spots		
CCC	0.61	0.07 (0.25)
CTC	0.64	0.26 (0.35)
GCC	0.05	0.29 (0.28)
GTC	0.28	0.41 (0.16)
SYC average	0.40 ± 0.25	0.26 ± 0.10

The mutability index is defined as the number of times a given oligonucleotide within a segment of DNA contains a mutation, divided by the number of times the oligonucleotide would be expected to be mutated for a mechanism with no sequence bias²⁰. The *in vitro* mutability index values were calculated from the spectrum in Fig. 2, and the *in vivo* data for trinucleotides 5'-NNC and also for complementary 5'-GNN (in parentheses) were taken from ref. 20. W = A or T; S = G or C; R = purine; Y = pyrimidine.

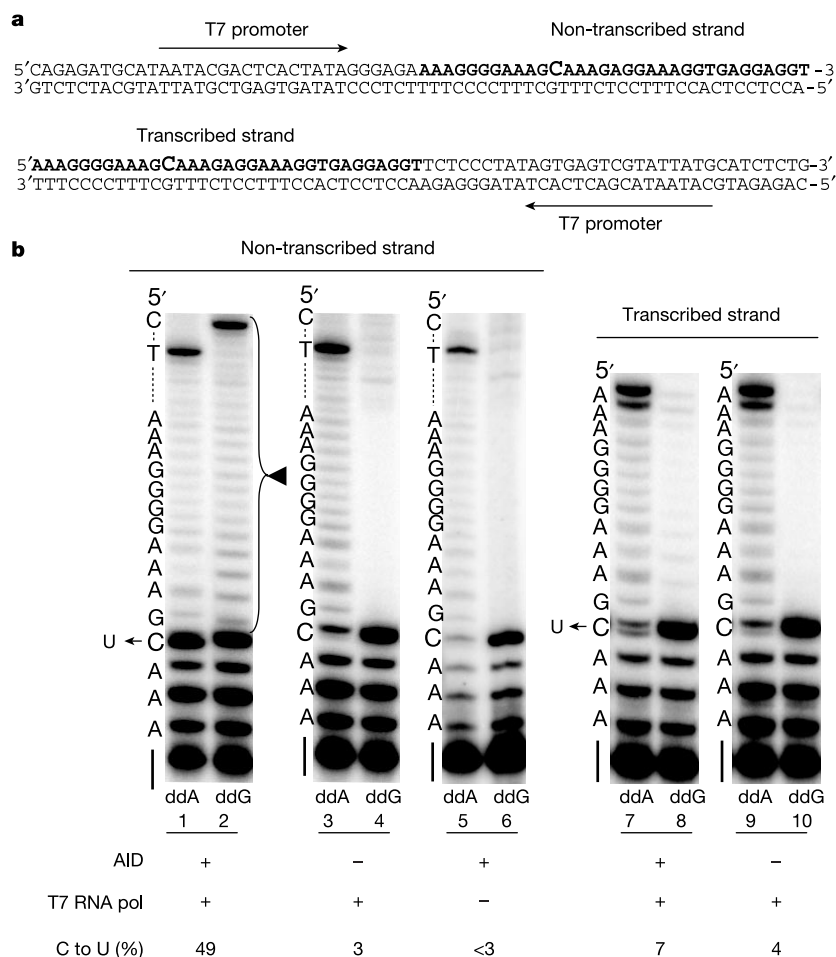


Figure 3 AID preferentially deaminates C in the non-transcribed strand during *in vitro* transcription of dsDNA. **a**, dsDNA sequences used for transcription by T7 RNA polymerase. The target sequence used to measure C deamination (bold) is distinct from the T7 promoter sequence. Target C residues located on either the non-transcribed (top) or transcribed (bottom) strands are indicated in bold in a larger font size. **b**, Polyacrylamide gel electrophoresis data showing AID-catalysed C → U deamination on the non-transcribed DNA strand dependent on the presence of T7 RNA polymerase (left panels), and a significantly (about 15-fold) smaller amount of deamination on the transcribed DNA strand (right panels). A primer elongation dideoxynucleotide termination assay² (see Methods) was used to measure the conversion of C → U, where the extent of deamination is measured by the presence of an intense termination band with ddA (lane 1) opposite U derived from the original template C site (U ← C, see sequence to the left of

gel). Concomitant with the conversion of C to U is the presence of bands migrating past the U ← C template site (ddG, lane 2), denoted by the bracket to the right of the gel. The absence of C deamination is observed by the absence of bands migrating beyond the U ← C template site in the ddG reactions (lanes 4, 6, 8 and 10). The lower and upper doublet bands at the U ← C position represent the presence of U and C, respectively (for example, lanes 7 and 9). The lower band of the doublet corresponding to U in lane 7 indicates that AID-catalysed deamination on the transcribed strand is about 15-fold less compared to the non-transcribed strand (after subtracting the background 4% in the absence of AID, deduced from lanes 9 and 10 for the transcribed strand). The three strong bands opposite A, at the bottom of the gels, are pause bands generated by Thermo Sequenase.

favourably with the corresponding C mutability index values found *in vivo* for SHM (Table 2). However, AID can only account for the most common C mutations, C → T transitions. The less common C → A and G transversions cannot arise by C deamination alone.

Active transcription is a prerequisite for SHM^{7,8}. As AID is inactive on dsDNA², transcription would have an essential role in SHM by generating regions of ssDNA on which AID can then act^{2,3,21}. Incubation of dsDNA with RNA polymerase in an *in vitro* transcription system including AID results in C deamination predominantly on the non-transcribed strand exposed by RNA polymerase (Fig. 3). Using a 'primer elongation dideoxynucleotide termination assay'² (see Methods), AID-catalysed C deamination is detected by the presence of a relatively intense termination band in the ddA lane, at the U ← C site indicated (Fig. 3 lane 1), and the presence of bands migrating beyond this site in the corresponding ddG lane (Fig. 3 lane 2; bracket shown to the right of the gel). A lack

of C deamination, observed as an absence of bands migrating past the U ← C site, demonstrates that deamination on the non-transcribed strand requires both AID (Fig. 3, lane 4) and RNA polymerase; that is, active transcription (Fig. 3, lane 6). Deamination on the transcribed strand, 'protected' as an RNA–DNA hybrid, occurs at an approximately 15-fold lower rate (Fig. 3, lanes 7 and 8, after subtracting background from lanes 9 and 10). These results agree with recently published data in *E. coli*²¹ and activated B-cell extracts³ showing that the mutations on the non-transcribed strand are dependent on transcription. Preliminary data with a transcription-dependent deamination assay containing a *lacZa* reporter sequence (R.B., P.P. and M.F.G., unpublished data), analogous to the experiment shown in Fig. 1, also demonstrate favoured deaminations in WRC SHM hotspot motifs and clonal heterogeneity in agreement with AID-catalysed C deamination activity on naked ssDNA (Table 1 and Fig. 1).

The assays described here provide new information relating the biochemical properties of AID to a number of salient aspects of SHM. Most notably, the preferential deamination of C by AID in SHM hotspots bears relevance to immunoglobulin V-gene mutational spectra, since AID also generates multiple deamination events on individual ssDNA molecules. The action of AID on ssDNA *in vitro* is consistent with the occurrence of multiple mutations and clusters of mutations in hypermutated clones^{22–24}, a poorly understood phenomenon. In this regard, the behaviour of AID represents a new and unanticipated result.

A perplexing property of AID purified from insect cells was its inability to catalyse C deamination on nucleic acid substrates owing to the presence of a tightly bound RNA inhibitor molecule². Whether or not inhibitory RNA bound to AID has a biologically important role or occurs adventitiously during purification, the avid binding of AID to single-stranded nucleic acids, owing to its overall positive charge, probably contributes to its non-distributive deamination pattern on ssDNA. Absent from our analysis are specific roles for error-prone polymerases^{25,26} and mismatch repair proteins²⁷, which will have to be accounted for in biochemical detail to provide a thorough understanding of SHM. However, the observation of AID-induced hypermutation in *E. coli* lacking ancillary eukaryotic factors¹³, taken alongside our observation that AID acting alone on a naked ssDNA substrate can distinguish between hot and 'not-so-hot' C sites while engaged in a process often causing multiple mutations, suggests that the specificity for immunoglobulin V-gene targeting at 5'-WRC hotspots may depend primarily, if not exclusively, on AID. □

Methods

Materials

M13mp2 phage and *E. coli* CSH50 were from our laboratory stocks. *Escherichia coli* strains MC1061 and its Ung[−] derivative were provided by R. Schaaper. AID was purified as a fusion protein, with glutathione S-transferase (GST) fused to the N terminus, as described previously² (see also Supplementary Information). T7 RNA polymerase and Thermo Sequenase were purchased from Promega and USB, respectively. M13mp2 gapped substrate DNA, with the gap region spanning positions −216 to +146 within the *lacZa* gene, was prepared as described previously¹⁵.

Mutation analysis of deamination on ssDNA *in vitro*

The specificity of AID-catalysed deamination of C to U was measured using a partially dsDNA M13mp2 circular substrate containing a 365-nucleotide ssDNA gap in the *lacZa* region. AID acts only on ssDNA, thus confining deamination events to the non-essential region of the M13 phage enabling detection of C → U conversion using the α-complementation assay¹⁵. The occurrence of C deamination within the gapped region gives rise to either white or light-blue M13 plaques. The plaques are dark blue in the absence of deamination. The deamination reactions (30 μl total volume), containing GST–AID (200 ng = 4 pmol), RNase (1 μg) and gapped DNA substrate (500 ng = 100 fmol) dissolved in a reaction buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1 mM dithiothreitol), were carried out at 37 °C for 20 min and terminated by twice extracting the DNA product with phenol:chloroform:isoamyl alcohol (25:24:1). The DNA products were transfected into competent MC1061 (Ung⁺) or NR9404 (Ung[−]) *E. coli* and plated on α-complementation host cells (*E. coli* CSH50) in the presence of X-gal and isopropyl-β-D-thiogalactopyranoside as described¹⁵. ssDNA from light-blue and colourless M13 phage plaques (arising from AID-catalysed C → U deamination), obtained from transfected Ung[−] cells, was isolated, and the entire gapped region was sequenced. In Ung⁺ cells, uracil residues are removed by uracil glycosylase followed by cleavage of the abasic sites by AP endonuclease, resulting in biologically inactive M13 DNA.

Transcription-dependent AID-deamination reactions

In a typical reaction (30 μl), RNase-activated GST–AID attached to glutathione–Sephadex beads² (10 μl), dsDNA substrate (30 nM) and T7 RNA polymerase (1 μl) in a reaction buffer (50 mM HEPES, pH 7.5, 1 mM EDTA, 10 mM MgCl₂) containing rNTPs (250 μM) were incubated at 37 °C for 30 min. RNase A (1 μg) was then added and incubated at 37 °C for 5 min to digest the nascent RNA to prevent interference with annealing of the primer during the primer elongation dideoxynucleotide termination assay (see below). The deamination reaction products were extracted twice with 40 μl phenol:chloroform:isoamyl alcohol (25:24:1). The extracted products were analysed using a primer elongation dideoxynucleotide termination assay, modified, as described below, from an assay reported previously². Thermo Sequenase was used to extend an 18-nucleotide ³²P-labelled primer annealed to the target strand in the presence of three dNTPs and either ddATP or ddGTP (for example see Fig. 3 dda or ddG lanes). The reaction was carried out for six cycles (95 °C for 30 s, 55 °C for 45 s, 72 °C for 5 min) and terminated by

adding an equal volume of stop solution containing 95% formamide and 20 mM EDTA. The cycling steps reduced the background on this long dsDNA template by carrying out the polymerization at a higher temperature to prevent nonspecific annealing of the primer. The reaction products were resolved by 20% polyacrylamide denaturing gel electrophoresis and analysed by phosphorimaging. Deamination efficiencies were calculated from extension reactions with the dda mix as a ratio of the band intensity opposite the C to U template site compared with integrated band intensities opposite and beyond this site. The efficiencies were also calculated from extension reactions with ddG as a ratio of integrated band intensities beyond the C template site to the integrated band intensities opposite and beyond the C template site.

Received 10 April; accepted 28 May 2003; doi:10.1038/nature01760.

Published online 18 June 2003.

- Kinoshita, K. & Honjo, T. Unique and unprecedented recombination mechanisms in class switching. *Curr. Opin. Immunol.* **12**, 195–198 (2000).
- Bransteitter, R., Pham, P., Scharff, M. D. & Goodman, M. F. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl Acad. Sci. USA* **100**, 4102–4107 (2003).
- Chaudhuri, J. *et al.* Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* **421**, 726–730 (2003).
- Dickerson, S. K., Market, E., Besmer, E. & Papavasiliou, F. N. AID mediates hypermutation by deaminating single stranded DNA. *J. Exp. Med.* **197**, 1291–1296 (2003).
- Liu, Y. J. *et al.* Normal human IgD⁺IgM[−] germinal center B cells can express up to 80 mutations in the variable region of their IgD transcripts. *Immunity* **4**, 603–613 (1996).
- Rada, C. *et al.* Immunoglobulin isotype switching is inhibited and somatic hypermutation perturbed in UNG-deficient mice. *Curr. Biol.* **12**, 1748–1755 (2002).
- Maizels, N. Somatic hypermutation: how many mechanisms diversify V region sequences? *Cell* **83**, 9–12 (1995).
- Peters, A. & Storb, U. Somatic hypermutation of immunoglobulin genes is linked to transcription initiation. *Immunity* **4**, 57–65 (1996).
- Muramatsu, M. *et al.* Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* **102**, 553–563 (2000).
- Revy, P. *et al.* Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* **102**, 565–575 (2000).
- Gerber, A. P. & Keller, W. RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem. Sci.* **26**, 376–384 (2001).
- Muramatsu, M. *et al.* Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J. Biol. Chem.* **274**, 18470–18476 (1999).
- Petersen-Mahrt, S. K., Harris, R. S. & Neuberger, M. S. AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* **418**, 99–103 (2002).
- Di Noia, J. & Neuberger, M. S. Altering the pathway of immunoglobulin hypermutation by inhibiting uracil-DNA glycosylase. *Nature* **419**, 43–48 (2002).
- Bebenek, K. & Kunkel, T. A. Analyzing the fidelity of DNA polymerases. *Methods Enzymol.* **262**, 217–232 (1995).
- Rogozin, I. B. & Kolchanov, N. A. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim. Biophys. Acta* **1171**, 11–18 (1992).
- Berek, C. & Milstein, C. The dynamic nature of the antibody repertoire. *Immunol. Rev.* **105**, 5–26 (1988).
- Dorner, T., Foster, S. J., Brezinschek, H.-P. & Lipsky, P. E. Analysis of the targeting of the hypermutational machinery and the impact of subsequent selection on the distribution of nucleotide changes in human V_HDJ_H rearrangements. *Immunol. Rev.* **162**, 161–171 (1998).
- Rogozin, I. B., Pavlov, Y. I., Bebenek, K., Matsuda, T. & Kunkel, T. A. Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. *Nature Immunol.* **2**, 530–536 (2001).
- Shapiro, G. S., Aviszus, K., Murphy, J. & Wysocki, L. J. Evolution of Ig DNA sequence to target specific base positions within codons for somatic hypermutation. *J. Immunol.* **168**, 2302–2306 (2002).
- Ramiro, A. R., Stavropoulos, P., Jankovic, M. & Nussenzweig, M. C. Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the non-template strand. *Nature Immunol.* **4**, 452–456 (2003).
- Gearhart, P. J. & Bogenhagen, D. F. Clusters of point mutations are found exclusively around rearranged antibody variable genes. *Proc. Natl Acad. Sci. USA* **80**, 3439–3443 (1983).
- Betz, A. G., Rada, C., Pannell, R., Milstein, C. & Neuberger, M. S. Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: Clustering, polarity, and specific hot spots. *Proc. Natl Acad. Sci. USA* **90**, 2385–2388 (1993).
- Michael, N. *et al.* Effects of sequence and structure on the hypermutability of immunoglobulin genes. *Immunity* **16**, 123–134 (2002).
- Goodman, M. F. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* **71**, 17–50 (2002).
- Kunkel, T. A., Pavlov, Y. I. & Bebenek, K. Functions of human DNA polymerases η, κ, and ι suggested by their properties, including fidelity with undamaged DNA templates. *DNA Repair* **2**, 135–149 (2003).
- Martin, A. & Scharff, M. D. AID and mismatch repair in antibody diversification. *Nature Rev. Immunol.* **2**, 605–614 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by National Institutes of Health Grants. P.P. and R.B. were supported on an NIH-NIA training grant. The authors acknowledge the intellectual contributions and encouragement of M. D. Scharff regarding all aspects of this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.F.G. (mgoodman@usc.edu).

Of mice and Tecnomouse

A model for oestrogen receptor defects, and *in vitro* antibody production.



Up for the cup: Stericup goes Express.

ERKO- β

Taonic www.taonic.com

Fertile area of research

The Taonic ERKO- β (oestrogen receptor- β) targeted mutation mouse model can be used to examine the role of oestrogen receptors in both reproductive and non-reproductive development. In this model, the *esr2* gene, which encodes a protein of about 60kDa, is disrupted in the embryonic stem cells, but *esr1*, the α -oestrogen receptor, is not and remains functional. Both sexes of the homozygous ERKO- β mouse develop normally and live to adulthood, but females experience infertility and have fewer offspring. The males have normal fertility, but suffer from prostatic and urinary bladder hyperplasia on ageing.

Gel-Pro Analyzer 4.5

Media Cybernetics www.mediacy.com

Software that reads between the lines

Gel-Pro Analyzer is designed for molecular biologists needing to obtain images from electrophoretic gels, blots and colonies and extract qualitative and quantitative information from them. Gel-Pro eliminates the need for manual input by automating molecular weight and amount calculation through the use of pattern recognition of objects, lanes and bands. This upgrade to version 4.0 also adds support for GLP/CFR 21 Part 11 compliance.

Express PLUS

Millipore www.millipore.com

Go with the flow

Millipore's Express PLUS membrane is now available in Stericup filter cups. These are fast-

flowing, low-binding cups for sterilizing tissue culture media, buffers and other aqueous solutions. The membrane is claimed to be faster than other 0.2 μ m polyethersulphone membranes, with surface chemistry designed to ensure that key growth factors and proteins are not absorbed. The threads of the bottle attachment are slightly recessed to prevent contamination. The cups are available in sizes to process volumes from 150 ml to 1 litre.

siRNA products

Dharmacon/Upstate www.upstate.com

Silent partners

Upstate has teamed up with Dharmacon to offer a new line of gene-specific small interfering RNA (siRNA) duplexes paired with qualified antibodies (siAB). RNA interference is an increasingly important method for functional genetic analysis and target validation. Each siRNA/siAB pair consists of a Dharmacon siRNA SMARTpool reagent (5 nmoles) and a corresponding Upstate validated antibody. SMARTpools consist of four siRNA duplexes

designed with Dharmacon's SMARTselection criteria against the target gene. siRNA SMARTpools are transfection-ready and are suitable to use for human gene knockdown.

Tecnomouse

Integra Biosciences

www.integra-biosciences.com

In vitro antibody production

Tecnomouse makes it possible to cultivate cells *in vitro* in a reliable and economic manner and to harvest the products over a period of months, minimizing the need for lab animals. The modular system is based on hollow-fibre membranes incorporated in a bioreactor known as the CultureCassette. Here, cell densities similar to those in tissue are produced. The compartmentalized design of the CultureCassette enables both O₂/CO₂ and media supply to be controlled and adapted to the specific needs of the cells.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Global Head of Naturejobs:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)

Netherlands/ Italy/ Iberia/ Belgium:

Evelina Rubio Hakansson (4973)
Scandinavia: Silje Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)
Natureevents:
Paul Constant (4954)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

International

Advertising Coordinator:
Hind Berrada (4935)

Naturejobs web development:
Tom Hancock

Naturejobs online production:
Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:
Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:
Rinoko Asami
e-mail: rasami@naturejpn.com

naturejobs

A demanding world

Multidisciplinary is officially yesterday's buzzword — cross-functionality is now the phrase to use. In a recent survey, New York-based recruitment consultancy TMP/Hudson found that demands in UK-based drug discovery and development companies have changed. Now, employees need not only to be able to communicate with researchers from different scientific backgrounds, they must also do a better job of working with their own company's range of commercial components.

TMP/Hudson interviewed 105 companies in the drug/biotech sector across Britain. Of these it found that only 41% believed that their employees had sufficient commercial acumen to complement their technical skills. This deficit is notable because companies are increasingly cautious about hiring in today's economic climate. Scientific skills and a multidisciplinary approach are now the default settings. The ability to understand how research can benefit a company's bottom line, then work with, say, business development or marketing to move a basic project into a commercial entity is at a premium, says Peter Rees, TMP/Hudson's main healthcare consultant in London.

And in the current climate, companies can demand such skills. In the past six months, only 44% of the firms interviewed had increased their staffing levels, and only a third planned to add more staff in the next six months. "They're planning growth, but they're being cautious," says Rees.

With that in mind, scientists looking to break into the drug discovery and development market might think about expanding their skill sets. 'Alternative' careers such as law, finance and marketing may increasingly be seen as necessary ancillary abilities. Until, that is, the job market opens up and encourages a new era of specialization — and the need for another buzzword.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

Benefiting as a
part-time teacher

p110

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS



For researchers in industry, a part-time teaching job can prove to be hugely beneficial. Myrna Watanabe reports from the classroom.

Leading a double life may not be for everyone, but for many researchers who are pursuing their career in industry, it can allow them to stay in touch with their roots. By becoming adjunct professors — part-time instructors at US colleges and universities — these researchers have found that they can keep a foot in academia and have their scientific assumptions refreshingly challenged by a stream of new students.

If the figures are anything to go by, the idea is a success. According to the American Association of University Professors, the proportion of faculty members who are adjunct professors has more than doubled to 43% in the past 20 years. One of the factors driving this shift is simple economics — increased enrolments at state universities and community colleges, which are suffering budget problems, mean they must either turn students away or hire adjuncts.

But there has also been a marked change in the

nature of the adjuncts themselves. No longer are they just people cobbling together two, three or more part-time teaching jobs, usually with no benefits, to make a living wage. Instead, many tend to be scientists with a full-time job in either research institutions or companies. And although some become adjuncts through necessity (see 'Multiple choices', below), many are ignoring the fact that the salaries tend to be disproportionately lower than for a full-time faculty job and seem to have much more altruistic motives.

According to the US National Science Foundation in 2001, nearly 41% of science PhDs doing research outside a university setting had a secondary job as a teacher. The main reason for this was simple: they loved teaching. Some liked it so much, they did it for free in addition to their full-time jobs. Some also saw teaching as benefiting them as well as their students, by feeding new ideas and potential recruits back to their full-time employers.

Take John Quackenbush. Full-time, he is a researcher at The Institute for Genomic Research (TIGR) in Rockville, Maryland, where he works on functional genomics and bioinformatics. But he also teaches at the George Washington University (GWU) in Washington DC and will soon be giving classes at Johns Hopkins University, where he will be involved in the master's course on bioinformatics and genomics that he helped to set up.

The job at George Washington is part of an agreement between the university and TIGR, under which Quackenbush gives lectures for graduate courses in genomics and participates in a graduate seminar. He is not paid for his teaching, but Quackenbush gets GWU graduate students to work in his lab and the university gets access to TIGR's genomic expertise.

Normally, adjunct teachers are paid \$1,500–4,500 per course, but for some, the money can actually be a negative



Extra work: John Quackenbush has set up genomics courses.

Multiple choices

Although today's adjunct professors tend to have full-time day jobs, some still take the traditional route of stringing together several part-time posts at different institutions to give themselves a full-time equivalent. Sometimes this is by choice, other times by necessity.

Helen Ortins Boland, for example, began her career in biotechnology. But she voluntarily left it 16 years ago to teach biology as an adjunct, and has been doing it ever since. She teaches three to four courses per semester. At one

point in her career, she taught at two different community colleges, travelling more than 60 kilometres between campuses. She says that she didn't mind because the flexible hours allowed her to raise her children. "I've missed very few soccer games," Ortins Boland says.

Claudia Kraemer is also gluing jobs together, but only while seeking a full-time appointment. She has a PhD in molecular biology from Italy, where, she says, adjuncts are unheard of. She teaches an average of 15 hours per week, five to six courses

per semester, at three different campuses. Fortunately, all the courses she teaches are interrelated, so she can write a basic lecture and adapt it to suit the course.

To help other adjuncts argue for benefits such as health insurance, pensions, paid sick days, and even use of office space or a telephone, Kraemer has become active on one community college's adjunct faculty advisory board. That activity, hopefully, will benefit all adjuncts no matter what their motivation for teaching.

M.W.



factor. "It ends up kind of disrupting your income tax at the end of the year," says microbiologist David Hodge. "I usually end up working a semester for free." Hodge works at the National Cancer Institute in Frederick, Maryland, but he also teaches a course in signal transduction for the Johns Hopkins biotech master's degree.

MONEY NO OBJECT

Marty Hall, who has been teaching on the part-time master's degree in computer science at Johns Hopkins for 16 years, agrees that most of the adjuncts teaching the degree aren't motivated by money. Hall runs a software training and consulting company called coreservlets.com in Owings Mills, Maryland. Although he is paid fairly well for an adjunct — he gets \$5,000 for a 14-week course that meets once a week for two hours and forty minutes — Hall and others note that there are downsides. "Sometimes, it does take time away from my family while I'm grading exams," Hodge says. Teaching does have other rewards, besides remuneration, such as helping Hodge to keep current in his field, that make such sacrifices worthwhile.

Civil engineer Jack Healy, who works full-time for Timken, a precision-bearing manufacturer in

Torrington, Connecticut, has been an adjunct teacher at what is now Naugatuck Valley Community College in Waterbury since 1989. He teaches environmental engineering and safety, which encompasses courses such as soil science, environmental measurements, waste minimization and hazardous materials. He finds teaching enriching because many of his students are his age or older, which provides him with insight and motivation to describe how theory is applied in real-world settings. He is so immersed in preparing his coursework that he now carries a camera with him when he travels for his job, so that he can take photos to use in his soil-science classes.

"I think teaching makes you a better scientist," says Hodge, who explains that the demands of his diverse classes force him to read materials outside his scientific research niche.

For Fredric Abramson, who has a PhD in human genetics, an MBA and a law degree, the students he taught at Johns Hopkins gave him a particular boost. "One of my students quipped: 'If you're so smart, why don't you make genomics useful?'" he says. That question sowed the seed that eventually led to him starting AlphaGenics, a biotechnology firm based in Gaithersburg, Maryland.

Barbara Finn, vice-president of regulatory affairs and quality assurance at drug-development firm Neurocrine Biosciences in San Diego, finds that her motivation for being an adjunct teacher comes from having an outlet beyond the company. Until last year, Finn taught for San Diego State University's masters on regulatory affairs. "I loved having an audience," she says.

The new breed of adjunct teachers are finding a number of reasons to invest their time in teaching. Whether it be keeping up-to-date with basic science, meeting potential recruits or collaborators, maintaining links with academia, or simply connecting with an audience, these adjuncts have learned that their part-time job makes a rewarding supplement to their full-time occupation. ■

Myrna Watanabe, is an adjunct professor and freelance writer in Patterson, New York.

Jack Healy has found acting as an adjunct professor to be inspiring.

Getting connected

With many employers downsizing, the number of people seeking adjunct positions seems to be increasing.

Adjunctopia, a website offering a free placement service for adjuncts, gets 30–50 new people signing up each day. Although the site currently lists few jobs, it has a database of 12,000 prospective adjuncts in all teaching fields.

Science seems to be a growth area for adjuncts

— especially in areas where industry has a need for more trained professionals, such as bioinformatics. "There are not a lot of people with degrees in these areas," says Joe Pensa, president and founder of Adjunctopia. He says that colleges want people with industry experience to teach professional courses, not people who have spent their careers isolated in academia. **M.W.** www.adjunctopia.com

